SNOMED Clinical Terms®
Technical Implementation Guide

January 2009 International Release

This document contains material useful to technical implementers working to incorporate SNOMED CT into software. However, it is subject to a continuing process of extension, quality review and revision.

Readers are advised to check any points of detail with the file structures of SNOMED CT as documented in the Technical Reference Guide. Similarly, anyb references to clinical standards or editorial policy within this document should be considered illustrative only; please refer to the SNOMED CT User Guide for documentation of clinical and editorial policies.

This document is intended to assist rather than direct or constrain developers. Implementers who make development decisions based on the advice herein do so entirely at their own risk.

The International Health Terminology Standards Development Organisation and the authors have applied their best endeavors to this document, but do not offer any warranty in respect of the accuracy or appropriateness of the information for implementation in any technical environment.

Table of Contents

Document History

Version	Notes
12-Oct-00	Outline posted for discussion
25-Oct-01	Initial sections from DM plus references to material from other contributors as agreed by TWG.
07-Jan-02	Most materials merged from other sources
22-Jan-02	Merged Migration document
31-Jan-02	SNOMED CT release draft
13-Feb-02	SNOMED CT First Release
01-Jul-02	SIEB 2002-06-13 agreed changes; support for moving concept between namespaces; additional status values; additional special concepts; more consistent descriptions of activity of concepts and descriptions in relation to component status; addition of information on Duplicate Terms Subset
26-Jul-02	• Updates to mapping services guide (Section 8)
July 2002	• Updates to Annex D4.4 on use of LATERALITY qualifier
January 2003	• Added inventory of documentation and updated copyright statements
July 2003	• Added inventory of documentation and updated copyright statements • Updated Cross References and Hyperlinks • Added Acknowledgements section • Added SNOMED CT Glossary • Added Distribution formats for the Practical Uses for Subsets sections from Subset Tables and Mechanisms technical specification document (v.19; 2001-12-12)
January 2004	• Release changes to reflect merger of Finding and Disease hierarchies
January 2005	• Merged Relationships and Historical Relationships files • “Part of” relationship no longer a defining relationship
July 2005	• Updates to Annex D, Annex F, and Glossary
January 2006	• Revised document formatting and headers • Updated SNOMED CT Glossary (Appendix G)
July 2007	• Updates to reflect transfer of IP to the International Health Terminology Standards Development Organisation • Removal of references to College of American Pathologists (CAP) derivative products
January 2008	• Updated for the January 2008 International Release
July 2008	Updated for the July 2008 International Release • Release of References Table • Updated diagrams and figures
January 2009	Updated for the SNOMED CT International Release

Inventory of Documentation

The following essential SNOMED CT documentation is currently available in both English and Spanish versions as part of the International Release of SNOMED CT from the International Health Terminology Standards Development Organisation (IHTSDO):

SNOMED CT Technical Reference Guide (TRG)

The TRG is intended for SNOMED CT implementers, such as software developers. The TRG assumes an information technology background. Clinical knowledge is not a prerequisite.

The TRG contains reference material related to the current release of SNOMED CT and includes file layouts, field sizes, required values and their meanings, and high-level data diagrams. It can be used to install and use SNOMED.

SNOMED CT Technical Implementation Guide (TIG)

The TIG is intended for SNOMED CT implementers, such as software designers. The TIG assumes information technology and software development experience. Clinical knowledge is not required, although some background is helpful to understand the application context and needs.

The TIG contains guidelines and advice about the design of applications using SNOMED CT, and covers topics such as terminology services, entering and storing information, and migration of legacy information.

SNOMED CT User Guide

The User Guide is intended for clinical personnel, business directors, software product managers, and project leaders; information technology experience, though not necessary, can be helpful.

The User Guide is intended to explain SNOMED CT's capabilities and uses from a content perspective. It explains the content and the principles used to model the terminology.

Additional Documentation

The following supplementary documentation is also included, in English only, as part of the International Release of SNOMED CT:

- SNOMED CT Canonical Table Guide
- SNOMED CT Developer Toolkit Guide
- SNOMED CT Namespace Identifier Guide
- SNOMED CT Namespace Registry

Guiding Principles, Development Process, and Acknowledgements

SNOMED CT: A Comprehensive Terminology for Health Care

In 1999, the College of American Pathologists (CAP) and the U.K. formed a strategic alliance to create a convergence of SNOMED® Reference Terminology (SNOMED® RT) and Clinical Terms Version 3 (CTV3). The resulting work, SNOMED Clinical Terms® (SNOMED CT®) combines the robust strength of SNOMED RT in the basic sciences and laboratory and specialty medicine with the primary care content of CTV3 (formerly known as the Read Codes). SNOMED CT is a comprehensive clinical reference terminology that provides clinical content and expressivity for clinical documentation and reporting. The terminology enables clinicians, researchers and patients to share comparable data. SNOMED CT was founded on four basic principles that have guided development activities related to the distribution table structure and clinical content:

- Development efforts encompass broad, inclusive involvement of diverse clinical groups and medical informatics experts.
- Clinical content is quality focused and adheres to editorial policies.
- A quality improvement process open to public scrutiny and vendor input, to ensure that the terminology is useful within healthcare applications.
- There should be minimal barriers to adoption and use.

The design has been driven by the expressed needs of software developers for features that improve their ability to develop useful applications.

SNOMED CT provides a standardized clinical terminology that is essential for effective collection of clinical data, its retrieval, aggregation and re-use, as well as interoperability.

SNOMED CT quality development process

The SNOMED CT development process incorporates the efforts of internal and external modelers. Content is edited by clinical editors who follow formal modeling guidelines. The integration of SNOMED RT and Clinical Terms Version 3 to create the first release was a three-year process that involved several stages of review and quality assurance:

- **Description mapping:** NHS editors evaluated each SNOMED concept and term and mapped it to the Clinical Terms Version 3 terminology; SNOMED editors performed the same task mapping primarily disorders and procedures from Clinical Terms Version 3 to SNOMED RT.
- **Description mapping conflict resolution:** Mapping discrepancies that occurred between NHS and SNOMED editors underwent a conflict resolution process to definitively place each concept within the merged hierarchy.
- **Autoclassification:** The merged database, following description mapping conflict resolution, underwent a series of quality control checks including autoclassification to identify and eliminate cycle errors (e.g. Concept A “IS_A” B and Concept B “IS_A” A) and equivalency errors (where two defined concepts have the exact same definition).
- **Ongoing refinement:** The quality control process is supplemented by feedback from users involved in adoption of SNOMED Clinical Terms. Parallel to domain specialist review, editors continue to review and edit content as needed.

Extent of review

The quality processes used in the development of SNOMED CT were complemented with external review.

- **Technical review:** The technical specifications for SNOMED CT were published for comment on both the SNOMED and NHS websites.
- **Alpha test review:** Forty two organizations in six countries tested the SNOMED CT alpha test file and completed a structured assessment instrument.
- **Alpha test feedback:** Debriefing sessions were conducted in the U.S., in the U.K. and in Australia, at which time test sites shared their positive experiences and recommendations for improvement.

Early adopters of SNOMED RT were debriefed on their implementation experience in order to identify the key issues to be addressed in the SNOMED CT Technical Implementation Guide.

Continuous quality improvement

Quality improvement is focused on updating the breadth and scope of the content to reflect changes in clinical care and advances in medical science; refining the content to deliver greater precision for data collection, retrieval and aggregation; and enhancing the functionality to serve users better.

Acknowledgements

Contributors to SNOMED CT®

SNOMED CT was originally created by The College of American Pathologists.

SNOMED CT has been created by combining SNOMED RT and a computer-based nomenclature and classification known as Clinical Terms Version 3, formerly known as the Read Codes Version 3, which was created on behalf of the U.K. Department of Health and is Crown copyright.

The IHTSDO also acknowledges the contributions of:

- The American Academy of Ophthalmology, for the ophthalmology-related portions of this work.
- SNODENT®: the Systematized Nomenclature of Dentistry, copyright 1998, American Dental Association. Used with permission.
- SNOVET®: the Systematized Nomenclature of Veterinary Medicine, copyright 1982, 1993, American Veterinary Medical Association. Used with permission.
- LOINC®: the Logical Observation Identifier Names and Codes, copyright 1995-2008, Regenstrief Institute and the LOINC Committee. All rights reserved.
- NANDA®: North American Nursign Diagnosis Association Taxonomy II, copyright 2005-2008, NANDA International. All rights reserved.
- The Perioperative Nursing Data Set® (PNDS), copyright 2002, AORN, Inc. All rights reserved.
- The Omaha System, copyright 1992, Martin and Associates. Used with permission.
- The Clinical Care Classification, copyright 2004, V.K. Saba. Used with permission.
- The Nursing Interventions Classification (NIC), copyright 2004, Mosby, Inc., and the Center for Nursing Classification and Clinical Effectiveness at the University of Iowa College of Nursing. Used with permission.
- The Nursing Outcomes Classification (NOC), copyright 2004, Mosby, Inc., and the Center for Nursing Classification and Clinical Effectiveness at the University of Iowa College of Nursing. Used with permission.

- This work contains material from the *AJCC Cancer Staging Manual, Sixth Edition* (2002) published by Springer-Verlag New York. Used with permission of the American Joint Committee on Cancer (AJCC), Chicago, Illinois
- The Anesthesia Patient Safety Foundation's (APSF) Data Dictionary Task Force. Some material contributed. Copyright 2003, APSF, Inc. Used by permission of the APSF.
- This work contains terms from the British Association of Dermatology (BAD), and is used by permission of BAD. Crown Copyright 2003 British Association of Dermatologists.
- This work contains terms from The Royal College of Anaesthetists (RCoA), and is used by permission of RCoA. Crown Copyright 2003 The Royal College of Anaesthetists.
- This work contains terms from the Authorized Osteopathic Thesaurus, and is used by permission of the American Association of Colleges of Osteopathic Medicine and the American Osteopathic Association.

1 Purpose

This document provides technical implementation guidance to assist in the effective development of SNOMED CT Enabled Applications.

2 Who Should Read This Document?

The guide can be used in various ways to assist the design or evaluation of various types of software applications that use SNOMED CT. The intended audience includes systems developers, health informatics specialists, purchasers, and system integrators.

2.1 Software developers

- Developers of fully integrated applications should use the guide:
  - As a checklist of SNOMED CT services necessary to meet the needs of their users.
  - For advice on how to implement the required services in ways that make the best use of SNOMED CT and which avoid known pitfalls.
- Developers of terminology servers should use the guide:
  - As a checklist when deciding which SNOMED CT services their server should offer
  - For advice on ways to implement the required services in ways that make the best use of SNOMED CT and which avoid known pitfalls.
  - As a point of reference when describing the functionality of their server.
- Developers of applications that use terminology services should use the guide:
  - As a checklist of SNOMED CT services necessary to meet the needs of their users.
  - To assist consideration of whether to use a terminology server.
  - As a point of reference when reviewing the functionality of terminology servers.

2.2 Health informatics specialists, analysts, purchasers and integrators

- Health informatics specialists analyzing the needs of users and organizations should use this guide:
  - As a checklist of SNOMED CT services necessary to meet the needs of their users.
  - For advice on known pitfalls when implementing clinical terminologies
  - To assist decisions on technical approaches to design and implementation of applications that use SNOMED CT.
- Purchasers of healthcare information systems should use this guide:
  - As a checklist when specifying procurement requirements for applications that use SNOMED CT.
  - As a starting point for the evaluation of the SNOMED CT related technical features of the available systems.

- Healthcare information systems integrators should use this guide:
  - As a checklist for confirming the claimed functionality of SNOMED CT Enabled Applications.
  - For advice on alternative approaches to integration of SNOMED CT related services into a wider information system.

3 Comments

Further information about SNOMED CT is available on the Internet at:

www.ihtsdo.org

Please send feedback by email to:

support@ihtsdo.org

or contact the International Health Terminology Standards Development Organisation at:

IHTSDO
Rued Langgaards Vej 7, 5te
2300 Copenhagen S
Denmark

Tel: +45 3644 8736
Fax: +45 4444 8736

4 Introduction

4.1 SNOMED Clinical Terms – Content

Important Notice

This guide describes services that should be provided by software applications that implement SNOMED Clinical Terms. The IHTSDO supplies content that can be loaded into these applications, but it does NOT supply any of the software itself.

The IHTSDO believes that the existing marketplace of software tools and applications will be best served if the IHTSDO itself promotes a level playing field where the SNOMED CT content is equally accessible to all vendors, suppliers and developers, and where the IHTSDO does not create healthcare software (such as EHR) that competes with them.

Note that this guide refers to files that are included with the International Release of SNOMED CT. Contact the IHTSDO for more information.

4.2 SNOMED Clinical Terms – applications and services

SNOMED Clinical Terms is a terminological resource that can serve many roles in healthcare software applications. User requirements vary according to the roles for which it is used. Healthcare software applications usually address a particular set of requirements associated with one or more clinical and/or business processes. Detailed requirements for integrating SNOMED CT into a particular application inevitably depend upon intended uses, the perceptions of users and the available technical environments.

The following examples illustrate a few possible types of implementation:

- A SNOMED CT enabled clinical record system incorporating clinical data entry, decision support, links to knowledge bases, sophisticated analysis, order-report message interfaces, support for record communication or sharing, etc.
- A data warehouse storing and analyzing records expressed with SNOMED CT encoded concepts.
- A diagnostic departmental system sending reports that include SNOMED CT encoded concepts to other systems.
- A hand-held data collection device used for input of a limited range of frequently used coded concepts.
- A decision support system using SNOMED CT concepts to represent guidelines and protocols for distribution to other systems.
- A system designed to enable the creation of queries for use in analysis of data held by various other systems, some of which contain SNOMED CT encoded data.
- A coding system mapping SNOMED CT encoded concepts (entered manually or read from an electronic record) to administrative groupings or classifications such as DRGs or ICD10.
- A system designed to support design and/or implementation of messages that convey specified information using a specified set of SNOMED CT concept identifiers.

Taking into account the variability in user requirements, this document sets out guidelines for effective use of SNOMED CT. Some of the issues discussed may not be relevant to applications that make relatively limited use of SNOMED CT. On the other hand, more sophisticated implementations may incorporate functionality that is not covered in this guide.

This document is not a statement of compliance requirements, but some of its recommendations may be restated in conformance profiles for certain types of software that use SNOMED CT.

4.3 Overview of the guide

The guide is divided into several sections dealing with different aspects of implementation:

- **Structure and content guide [Chapter 5]**
  - A summary of features of SNOMED CT that need to be understood by those developing or implementing SNOMED CT Enabled Applications.
- **Terminology services guide [Chapter 6]**
  - A guide to services required to access SNOMED CT terminological resources. These services include searching for terms, navigating concept hierarchies, applying subsets and utilizing the defining and qualifying relationships of concepts.
- **Record services guide [Chapter 7]**
  - A guide to services related to entering, storing, retrieving or communicating SNOMED CT encoded information. These services differ from terminology services, as they are concerned with access to instances of data encoded using SNOMED CT, rather than to the terminology resources provided by SNOMED.
  - Most record services are dependent on terminology services and these two sets of services may be fully integrated in an application. Alternatively, an application providing record services may use terminology services provided by a general-purpose terminology server.
- **Legacy information migration guide [Chapter 8]**
  - A guide to issues related to conversion or use of records, queries and protocols originally created using a terminology other than SNOMED CT.
  - The main focus of the guide is on migration from SNOMED RT, SNOMED International, Clinical Terms Version 3 and the Read Codes. However, the general advice provided is also applicable to migration from other terminologies.
- **Extension services guide [Chapter 9]**
  - A guide to additional services which some advanced users may require to allow them to create or maintain *Extensions* for use in a particular specialty or realm.
  - The most common of these requirements will be to support the creation and maintenance of specialized *Subsets*.

4.4 Application design issues

When designing or implementing a SNOMED CT Enabled Application, the first step is to assess the range of services necessary to meet user requirements. The next step is to decide whether all the services are to be delivered by integral parts of a single application or whether reusable components can be used for some services.

If a component-based approach is taken, a logical division can be made between services that depend on the design of the application data structures and services that can be generalized in ways that are independent of the way information is recorded, stored or retrieved. It is realistic to consider providing and/or utilizing the more generalized services through general-purpose application programming interfaces (APIs). The guide does not specify the technical detail of such an interface but the structure of the guide makes a clear distinction between generalizable terminology services (see Figure 4.1), and the data-dependent record services (see Figure 4.2).

Integral and component-based approaches offer a different profile of advantages, some of which are summarized below.

- A component-based approach offers the following advantages:
  - Rapid development of SNOMED CT related functionality, focused on meeting the requirements of users of a specific software application.
  - Simplifies future migration to enhanced or more cost-effective solutions by separately identifying reusable and replaceable modules.
  - Allows several applications used by a single organization to use a single terminology server. This has several advantages:
    - Reduction of maintenance and support cost associated with installing each release of SNOMED CT.
    - Guaranteed alignment of SNOMED CT releases between applications that share the server.
    - Consistency of the user interface and technical characteristics of different applications with respect to their access to SNOMED CT.
- A fully integrated approach offers the following advantages:
  - Independence of third party development.
  - Customized access to SNOMED CT tailored to the needs of particular application users.

The approach chosen depends on a careful consideration taking into account the cost and functionality of available components. Commercial and technical concerns about dependence on third-party components may be a valid reason for in-house development of all the required services. However, even where all the development is undertaken within a single organization, separation of terminology and record services into separate components may offer a more robust approach, allowing future extensibility and migration at lower cost.

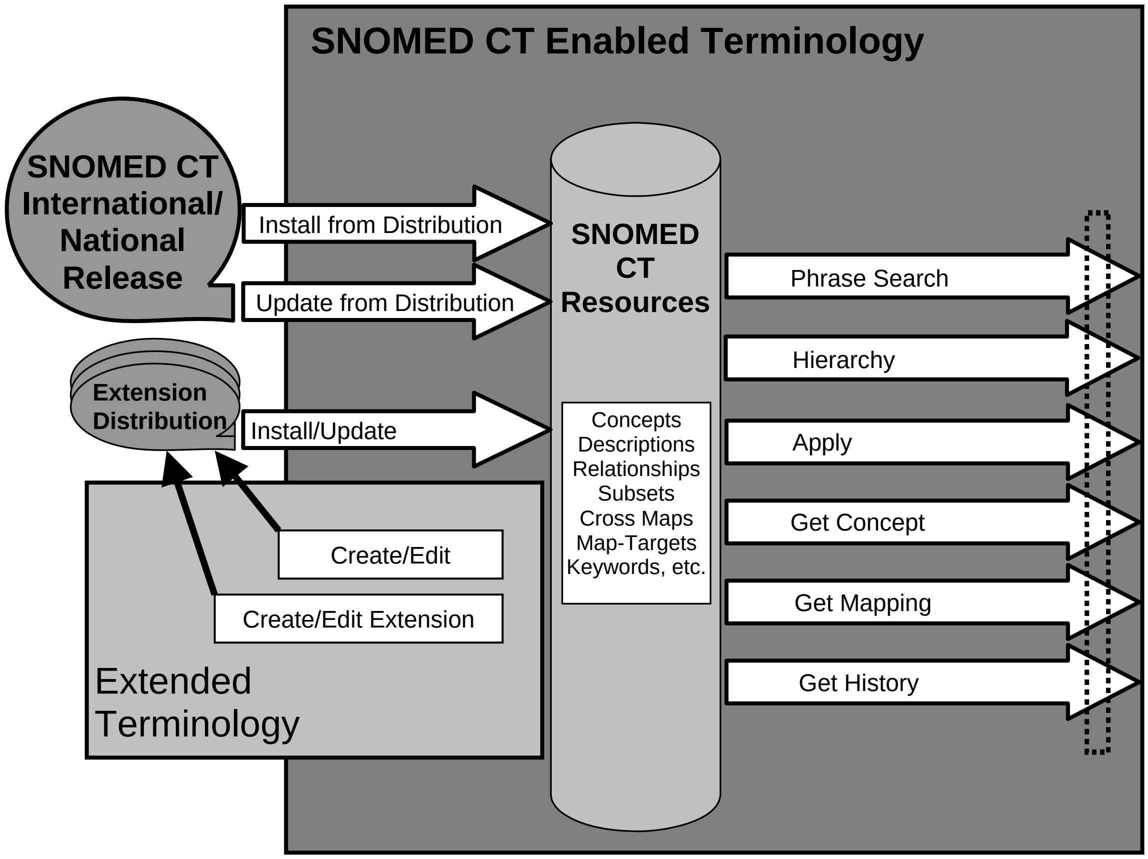


Figure 4.1: Overview of Terminology Services

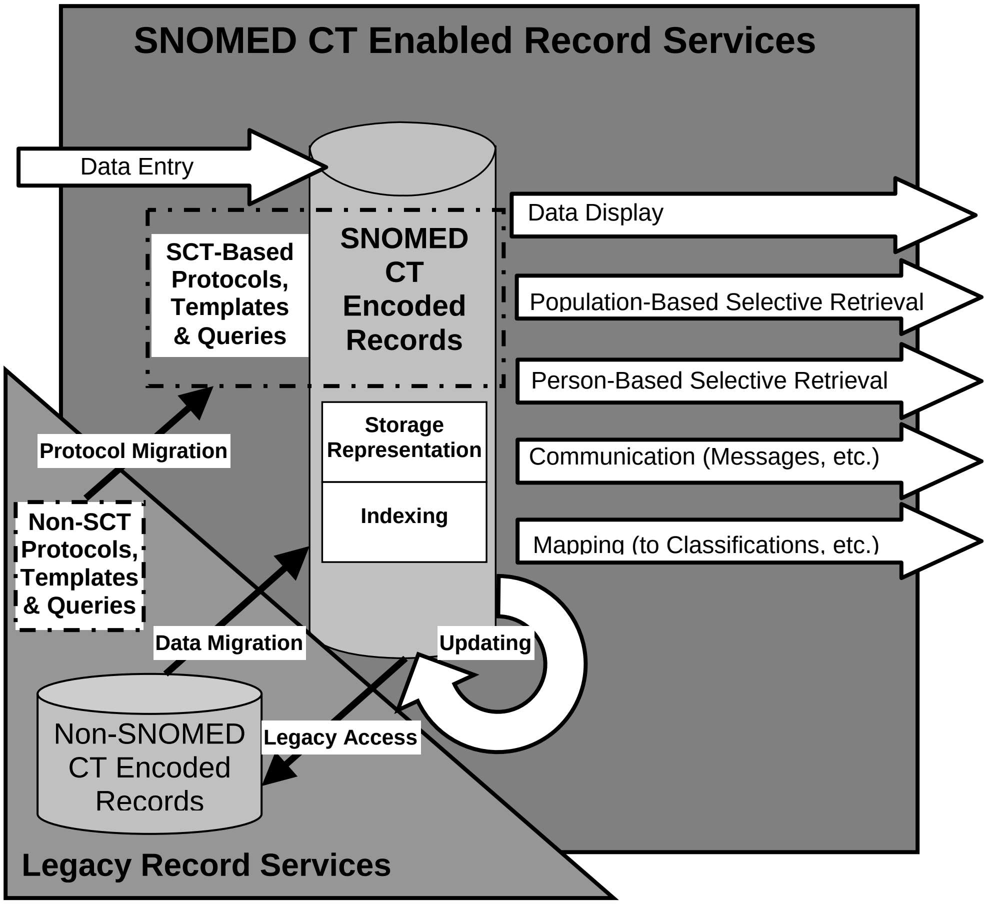


Figure 4.2: Overview of Record Services

5 Structure and Content Guide

5.1 Introduction

This section summarizes features of SNOMED CT that need to be understood by those implementing SNOMED CT Enabled Applications.

5.2 SNOMED CT Components

5.2.1 Concepts

SNOMED CT contains representations of hundreds of thousands of healthcare-related concepts.

Each Concept is identified by a unique ConceptID and is distributed as a row in the Concepts Table.

Each Concept includes alternative identifiers of the same Concept using:

- The five-character code used in Clinical Terms Version 3 (and earlier Read Codes versions).
- The six to eight-character code used in SNOMED International.

5.2.2 Descriptions

A Description associates a human-readable term with a Concept that it describes.

A Concept may be associated with multiple alternative Descriptions that represent the Preferred Term, Synonyms, or Fully Specified Name for the Concept in a particular language or dialect.

A Description may be a preferred name in one language and a synonym in another. This is indicated by references to the Description from an appropriate Language Subset.

Each Description is identified by a unique DescriptionID and is distributed as a row in the Descriptions Table.

5.2.3 Relationships

A Relationship represents an association between two Concepts.

Each Relationship is identified by a unique RelationshipID and is distributed as a row in the Relationships Table.

A Relationship contains identifiers of two logically associated Concepts and the identifier of another Concept that indicates the RelationshipType by which they are associated.

Example:

A Relationship may assert that "*arthritis*" (first related Concept) "IS_A" (relationship type) "*joint disorder*" (second related Concept).

A Relationship contains an indication of its CharacteristicType. This distinguishes between:

- Defining characteristics, which state things that are always true about a Concept.
- Qualifying characteristics, which offer options for qualifying a Concept.
- Context specific characteristics, which may vary according to place or time.

A Relationship has a Relationship group attribute which links interrelated defining characteristics.

Example:

"*Removal of a foreign body from the stomach by gastrotomy*" involves:

- "removal" of the "foreign body" (not to the stomach)
- "incision" of the "stomach" (not the foreign body)

A Relationship may itself be represented in a hierarchical structure called a role hierarchy.

Example:

Direct Device and Indirect Device are both subtypes in the Procedure Device role hierarchy.

- Concepts can have relationships defined with Direct Device, Indirect Device, or Procedure Device.
- Data retrieval can be constructed to recognize the role hierarchy membership and collect all Concepts with descendants of Procedure device.

5.2.4 Subsets

Subsets represent groups of Components that share specified characteristics that affect the way the components are displayed or otherwise accessible within a particular realm, specialty, application or context.

Different types of Subsets are used to represent:

- Language and dialect variants,
- Frequency of use of Descriptions or Concepts in particular realms or specialties,
- Suitability of particular Concepts for use in a particular context in a record,
- Structure and ordering of hierarchies displaying Concepts for user navigation.

Subsets are released in a relational form as a row in the Subsets Table and a set of rows in the Subset Members Table. The rows in the Subset Members Table contain a common SubsetID associating them with the appropriate row in the Subsets Table.

5.2.5 Cross Map sets

A Cross Map Set represents a set of mapping from SNOMED CT Concepts to one or more codes in a classification, coding scheme or administrative grouping.

The individual maps within a Cross Map Set are expressed as Cross Maps and Cross Map Targets.

Each Cross Map Set is identified by a unique CrossMapSetID and is distributed as a row in the Cross Map Sets Table.

A Cross Map maps a single Concept to a Cross Map Target. If there are several possible ways of mapping a particular Concept to the target scheme there are several Cross Maps associated with the same Concept. If there are several Cross Maps for a Concept, the Cross Maps may include text advice and or computer parsable rules indicating the conditions in which this map applies.

Each Cross Map contains an identifier of the Cross Map Set of which it is a member and is distributed as a row in the Cross Maps Table.

A Cross Map Target contains an expression that represents one or more mapped SNOMED CT Concept in the target scheme. A Cross Map Target may include more than one target code where this is necessary to express the SNOMED CT Concept in the target scheme. Depending on the nature of the Cross Map Set, a Cross Map Target may include text advice and/or computer parsable rules indicating the conditions in which this target applies.

Each Cross Map Target is identified by a TargetID and is distributed as a row in the Cross Map Targets Table.

5.2.6 Search support

Several tables are included in the Developer Toolkit to simplify support for text searching.

There are two WordKey Tables. These tables link each word used in SNOMED CT to every:

- Description in which it is used;
- Concept associated with a Description in which the word is used.

There are also two Dualkey Tables. These tables link each abbreviated word pair to every:

- Description in which that pair of words is used;
- Concept in which the combined set of Active Descriptions contains that pair of words.

These tables are provided to assist implementation. However, use of these tables is optional as developers may generate and use alternative search support resources.

An extended version of the Toolkit provides Java™ programs to generate indexes that may be useful to organizations that develop SNOMED CT Extensions.

5.2.7 Canonical forms

The Canonical Table is an additional release table that contains an alternative representation of the defining characteristics associated with each Concept. Use of this table is optional as it adds no new information but provides a pre-computed short canonical representation of each Concept.

The Canonical Table follows a structure similar to that of the Relationships Table but its content is somewhat sparser. It contains the minimal set of defining characteristics required to distinguish each Concept from other Concepts with the same primitive supertype ancestors.

The Canonical Table may be used to simplify some advanced retrieval and aggregation processing. However, use of this table is optional as the information it provides can also be computed from the released Relationships and Concepts.

5.2.8 History information

Once released, the unique identifiers of SNOMED CT components are persistent, and their identifiers are not reused.

Concepts and Descriptions continue to be distributed even when they are no longer recommended for active use. This allows a current release to be used to interpret data entered using an earlier release. The ConceptStatus or DescriptionStatus fields are used to indicate the reason for inactivating a Component. These reasons include errors, duplication of another component and ambiguity of meaning. Some SNOMED CT Concepts represent classification concepts that have imprecise and potentially changeable meanings. These remain active but are marked with the ConceptStatus "limited."[1]

Other Components (i.e. Relationships, Subsets, Cross Map Sets, Cross Maps and Cross Map Targets) are not distributed if they are no longer in active use.

When a Component is added, inactivated or otherwise changes its status, this is noted in the Component History Table.

When a Concept is no longer in active use, is replaced by or duplicates another Component, this is indicated by an appropriate historical relationship. When a Description, Subset or Cross Map Set is replaced by or duplicates another Component, this is indicated by a row in the References Table.

5.2.9 Component identifiers

The Concepts, Descriptions, Relationships, Subsets, Cross Maps, and Cross Map Targets described in the previous sections are all SNOMED CT Components.

Each SNOMED CT Component has a unique identifier. This identifier is an integer that can be expressed in a 64-bit binary form or as a character string consisting of between 6 and 18 digits. The specification of the identifier includes:

- A computable check-digit to reduce the risk of errors.
- A partition-identifier dependent on the types of Component.
    - Separate partitions are specified for Components originating from the SNOMED organization and those created as legitimate Extensions for use in a particular realm.
    - In the case of the extension partitions, a namespace-identifier is also specified to allow different organizations to maintain Components without reusing the same identifiers.

5.3 Extensions

5.3.1 Introduction

SNOMED CT supports the inclusion of Extensions to the core terminology managed by different organizations. An Extension may contain Components of the various types (i.e. Concepts, Descriptions, Relationships, Subsets, Cross Map Sets and Cross Map Targets etc.).

5.3.2 Content of Extensions

The Components in an Extension have the same structure as core Components of SNOMED CT.

SNOMED CT core release data is not dependent on availability of any Extension. However, Extensions are not "free-standing" but are dependent on the core release of SNOMED CT.

Some Extensions may also be dependent on another higher-level Extension but these dependencies must not be circular.

[1] Some Concepts derived from classifications such as ICD 10 include the abbreviations NOS (not otherwise specified) or NEC (not elsewhere classified). These are only valid in respect of a particular classification and change in their meaning if additional precisely defined codes are added to that part of the classification. Furthermore, a Concept that is not otherwise specified in ICD10 may well be more precisely represented by another SNOMED CT Concept and thus from a SNOMED CT perspective "otherwise classified."

- Concepts in an Extension
  - Must be subtype descendants from a core Concept.
    - This descent may be indirect through other Concepts in the Extension and/or Concepts in a higher-level Extension.
- Descriptions in an Extension
  - May apply to a core Concept, a Concept in the same Extension or a Concept in a higher-level Extension
- Relationships in an Extension
  - May define or qualify Concepts in the same Extension
  - These Relationships may refer to Concepts from the core or from a higher-level Extension as RelationshipTypes and/or ConceptID2.
- Subsets in an Extension
  - May include as its members Components from the core, Components from the same Extension, Components from a higher-level Extension or a mixture of Components drawn from these sources.
  - May refer to a Subset in the core or higher-level Extension as a Base Subset from which variations are specified in a Subset Definition File.
- Cross Map Sets in an Extension
  - May provide maps for Concepts from the core, Concepts from the same Extension, Concepts in a higher-level Extension or a mixture of Concepts some drawn from the core and some from these sources.

5.3.3 Extension namespaces

The Components of an Extension have identifiers which follow the same form as that used within the core. However, these identifiers include a partition-identifier indicating that the Component is part of an Extension and a namespace-identifier specific to the responsible organization.

Partition-identifiers and namespace-identifiers serve three roles:

- Prevention of identifier collision or reuse.
  - Organizations responsible for an Extension must only issue Components within their allocated namespace and must not reuse any identifier within that namespace once it has been issued.
- Indicating maintenance responsibility
  - No organization is permitted to make any change to a Component which has an identifier that is not within their allocated namespace. This avoids the risk of two organizations updating or changing the ConceptStatus of the same Concept.
- Indicating the source for information about a ConceptID
  - If information containing ConceptIDs from a particular namespace is received by an application that does have the relevant Extension installed, the application can check to determine if the source of the namespace is known.
  - The responsibility for an allocated namespace remains with the organization to which it was issued unless responsibility is transferred by merger or mutual agreement. Any namespace transfer must be notified to and authorized by the IHTSDO.

5.3.4 Transfer of responsibility for Components

In some situations it is appropriate to transfer responsibility for management of one or more Concepts between separately managed namespaces. These transfers may include transfers to and from the SNOMED CT core as well as transfers between Extensions. When responsibility for the maintenance of Concepts is transferred between organizations with different namespace identifiers, this is represented by status changes (Moved Elsewhere and Pending Move).

5.3.5 Namespace allocation

Namespace-identifiers are allocated by the IHTSDO to licensed organizations. The IHTSDO is under no obligation to allocate a namespace to any organization and makes these allocations at its discretion.

Allocation of a namespace does not imply any endorsement of the reputation of an organization nor to the quality or fitness for purpose of any Extensions created by that organization. Users and/or vendors incorporating Extensions into their application do so at their own risk and should satisfy themselves with the reputation of the responsible organization and the quality the Extensions so incorporated.

5.4 Essential Concepts and Relationships

5.4.1 Root Concept

Role of the root Concept

The root Concept has the Fully Specified Name "SNOMED CT Concept (SNOMED RT+CTV3)".

All other Concepts are subtypes of the root Concept.

The root Concept is not the source of any Relationships. This means that there are no Relationships in which the ConceptID1 value is equal to the ConceptID of the root Concept.

Representation of the root Concept

The root Concept has a designated ConceptID; this is documented in Appendix B.

Release information in the root Concept

The root Concept has a current Synonym that contains information about the release. The Synonyms, representing earlier releases, are distributed as Inactive Descriptions. The release information is represented in the term text of the Synonym as indicated in Table 5.1.

Table 5.1: Representation of release information in the root Concept

Example	SNOMED Clinical Terms version: 20020131 [R] (first release)	
Stylized form	SNOMED Clinical Terms version: yyyymmdd [status] (description)	
	yyyymmdd	The release date in ISO format.
	Status	R (release), D (developmental) or E (evaluation).
	Description	An optional free text description of the release.

5.4.2 Subtype Relationships

Role of subtype Relationships

Subtype Relationships provide the main semantic hierarchy that relates Concepts to one another.

All Concepts, except the root Concept, have subtype Relationships with one or more Concepts. Each of these Relationships indicates that a Concept is a subtype of another Concept.

Representation of Subtype Relationships

Subtype Relationships are expressed in the same way as all other SNOMED CT Relationships. They are identifiable by their RelationshipType, which refers to a Concept with the Fully Specified Name "IS_A."

The subtype Relationship Concept has a designated ConceptID, which is documented in Appendix B.

Subtype Relationships and the Subtype Hierarchy

Table 5.2 and Figure 5.1 provide two views of the same set of subtype Relationships. These illustrate the following features of subtype Relationships.

Only the most proximate subtype relationships are represented explicitly in the distributed Relationships Table. These Relationships are shown as blue lines in Figure 5.1.

A Concept is a subtype of:

- Concepts to which it has an explicit subtype Relationship.
  - Thus the Concept "bacterial pneumonia" is a subtype of "bacterial infectious disease" and is also a subtype of "pneumonia."
- Concepts that are indirectly linked via a sequence of one or more subtype Relationships.
  - Thus, the Concept "bacterial pneumonia" is also a subtype of "pneumonia," "disease of lung," "infectious disease" and, in fact, all the other Concepts in the example up to and including the root Concept.

Example:

"*Bacterial pneumonia*" is a subtype of "*pneumonia*" because it is a subtype of "*infective pneumonia*" which is a subtype of "*pneumonia*."

The number of links in the chain of subtype Relationships between two Concepts does not alter the logical semantics of the relationship between them. The number of subtype Relationships between two Concepts may change between releases of SNOMED CT as a result of the addition of an intermediate Concept. This does not alter the semantic relationship between them.

Some technical implementation issues are affected by whether a pair of Concepts is linked by a single subtype Relationship or by a sequence of several subtype Relationships. In this guide, the following terms are used where this distinction is technically significant:

A given Concept (Concept-x) may have:

- Subtype children[2] – Concepts with a subtype Relationship referring to Concept-x.
  - "*bacterial pneumonia*" is a subtype child of:
    - "*bacterial infectious disease*"
    - "*infective pneumonia*"
- Supertype parents[3] – Concepts referred to by a subtype Relationship from Concept-x.
  - "*infectious disease*" is a supertype parent of:
    - "*bacterial infectious disease*"
    - "*infective pneumonia*"
- Subtype descendants – Concepts with subtype Relationships that refer to other Concepts that are either child or subtype descendants of Concept-x.
  - "*bacterial pneumonia*" is a subtype descendant of:
    - All other Concepts shown in the example.
- Supertype ancestors – Concepts referred to by subtype Relationships from other Concepts that are either parent or supertype ancestors of Concept-x.
  - "*disease*" is an supertype ancestor of:
    - All other Concepts shown in the example, except for "SNOMED Clinical Terms Concept."

[2] Subtype children are also referred to as "Immediate subtypes."
[3] Supertype parents are also referred to as "Immediate supertypes."

Table 5.2: Tabular view of the "IS_A" relationships for an example Concept[4]

Supertypes of bacterial pneumonia					
bacterial pneumonia *IS_A* bacterial infectious disease					
	bacterial infectious disease *IS_A* infectious disease				
				infectious disease *IS_A* disease	
					disease *IS_A* SNOMED Clinical Terms Concept
bacterial pneumonia *IS_A* infective pneumonia					
	infective pneumonia *IS_ A* pneumonia				
		pneumonia *IS_A* disease of lung			
			disease of lung *IS_A* disease of respiratory system		
				disease of respiratory system *IS_A* disease	
					disease *IS_A* SNOMED Clinical Terms Concept
infective pneumonia *IS_A* infectious disease					
				infectious disease*IS_A* disease	
					disease *IS_A* SNOMED Clinical Terms Concept

[4] Note that the relationships shown in the table and diagram are not the definitive released Relationships of these Concepts. They have been simplified to illustrate particular points in the text.

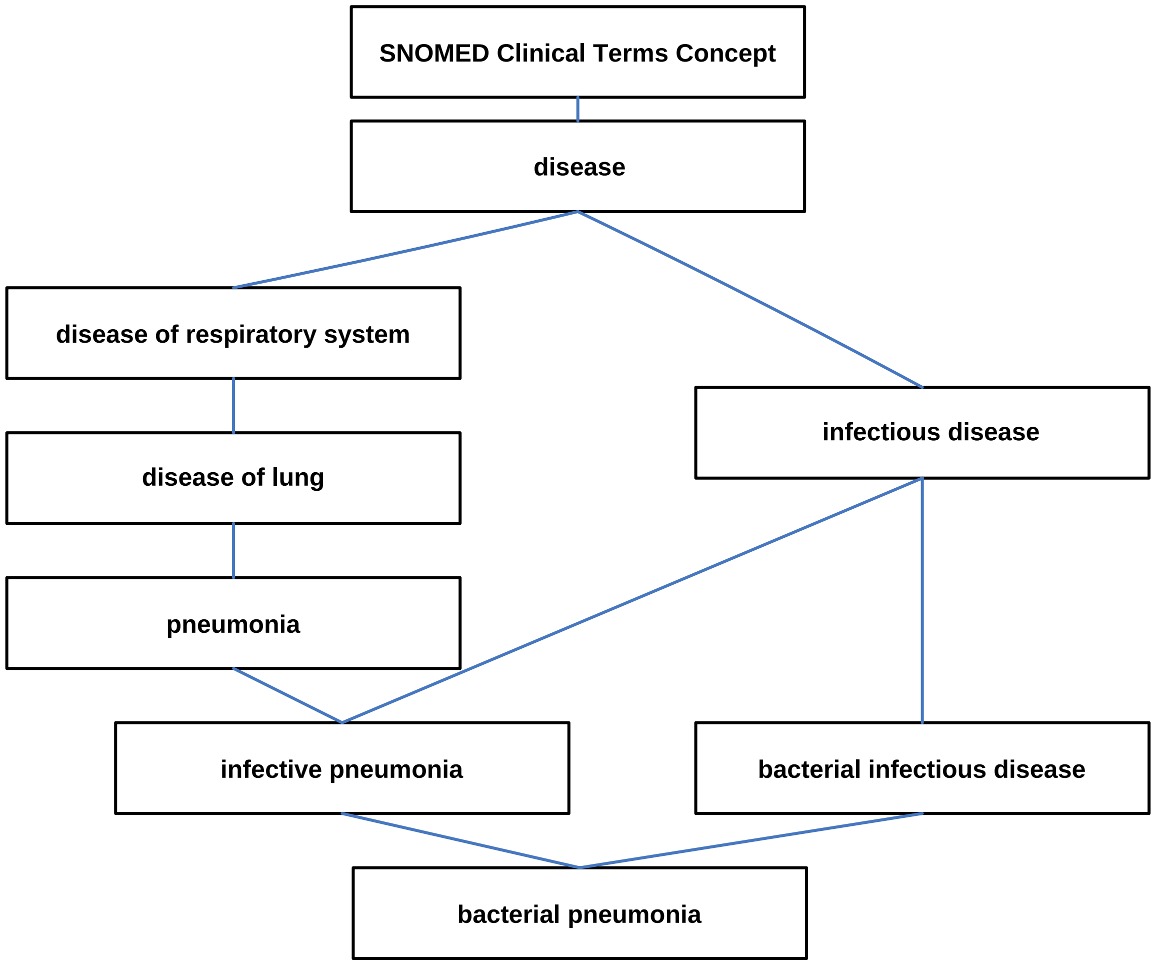


Figure 5.1: Graphical representation of the "IS_A" relationships of an example Concept

5.4.3 Top-level Concepts

Role of the Top-Level Concepts

The top-level of the subtype hierarchy contains Concepts that represent broad semantic types. These include:

- *Linkage concept*
- *Body structure*
- *Situation with explicit context*
- *Environment or geographical location*
- *Event*
- *Clinical finding*
- *Observable entity*
- *Organism*
- *Pharmaceutical/biologic product*
- *Physical force*
- *Physical object*
- *Procedure*
- *Qualifier value*
- *Record artifact*
- *Social context*
- *Specimen*
- *Staging and scales*
- *Substance*
- *Special concept*

As explained in the previous section, a Concept can have more than one supertype parent. However, each Concept is a subtype descendant of one and only one top-level Concept. Thus a Concept that is a "*disease*" cannot also be a "*procedure*."

In Figure 5.1, there are three distinct routes between the Concept "*bacterial pneumonia*" and the root Concept. However, all of these routes converge at or below the top-level Concept "*disease*."

Top-level Concepts with structural roles

Some top-level Concepts have specific structural roles in the terminology. These are indicated below:

- Linkage concept
  - All Concepts that can be used as Relationship Types are subtypes of the top-level Concept "Linkage concept."

- Qualifier value
  - Subtypes of the top-level Concept "Qualifier value" are only suitable for use as the value of an attribute. A qualifier value may be used in a defining Relationship and in others as a qualifier in a post-coordinated expression.
- Special concept
  - The subtypes of the top-level Concept "Special concept" are several types of Special Concepts that share a common characteristic but which are not part of the current logical hierarchy of semantic subtypes. These include:
    - Inactive concept – The supertype ancestor of all Inactive Concepts.
    - Navigation concept – The supertype parent of all Navigation Concepts.
    - Namespace concept – The supertype parent of all Namespace Concepts.

Representation of the top-level Concepts

Awareness of the top-level Concepts is likely to be particularly important when developing technical implementations.

A top-level Concept can be identified by the fact that it has a single subtype relationship referring to the Root Concept. However, to minimize processing requirements the top-level Concepts have designated ConceptIDs that are documented in Appendix B. Any subsequent additions or changes to these top-level Concepts will be clearly documented in release notes.

5.4.4 Defining characteristics

Role of defining characteristics

Defining characteristics provide the logical definitions of Concepts. The subtype relationships discussed in [5.4.2] contribute to the definition of each Concept. Additionally, Concepts may be further defined by additional attributes including etiology, topography, method, etc.

The range of attributes applicable depends on the type of Concept. For example, a procedure may have a method, and a disorder may have an etiology, but a procedure cannot have an etiology, and disorder cannot have a method.

Defining characteristics using a particular attribute will be applied consistently to all Concepts to which it is relevant. Note that this design principle may not be fully realized for all attributes in each release.

Representation of defining characteristics

Defining characteristics are represented as Relationships. The fields are used as follows:

- ConceptID1 – refers to the Concept to which a defining characteristic applies.
- RelationshipType – indicates the nature of the defining attribute
- ConceptID2 – refers to the Concept that represents the value of that attribute

In each release the supported defining characteristics for every Concept are distributed in the Relationships Table. The supported defining characteristics are descendents of the upper level Concept Attribute. Not all of the descendents of attributes are supported; the list of supported defining attributes is given in Table B.5 [Appendix B].

The Canonical Table contains only those defining characteristics that differ from the primitive supertypes of each Concept.

More information about the use of the Canonical Table and the Relationships Table is given in section [6.7.2] and in section [7.4.3], which deals with issues relating to data retrieval.

Table 5.3: Defining characteristics applied to an example Concept[5]

Disease		
	IS_A SNOMED CT Concept	
	Primitive	Not all SNOMED CT Concepts are diseases. No defining characteristics are included to specify what makes something a disease.
infectious disease		
	IS_A disease	
	causative agent infectious agent	
	Primitive	Not all diseases with causative agent "infectious agent" are "bacterial infectious disease". For example, "rheumatic heart disease has causative agent "streptococcus" but is not an "infectious disease."
bacterial infectious disease		
	IS_A infectious disease	
	causative agent bacteria	
	Fully-defined	All infectious diseases with causative agent "bacteria" are "bacterial infectious diseases"
disease of respiratory system		
	IS_A disease	
	finding site respiratory system structure	
	Fully-defined	All diseases with associated topography "respiratory system structure" are "diseases of respiratory system."
disease of lung		
	IS_A disease of respiratory system	
	finding site lung structure	
	Fully-defined	All diseases of respiratory system with associated topography "lung" are "diseases of lung."
Pneumonia		
	IS_A disease of lung	
	finding site lung structure	
	Primitive	Not all diseases of lung are "pneumonia." No additional characteristics specify what attributes are needed to specify "pneumonia."
infective pneumonia		
	IS_A infectious disease	
	IS_A pneumonia	
	causative agent infectious agent	
	finding site lung structure	
	Fully-defined	All pneumonias with causative agent "infectious agent" are "infective pneumonia."
bacterial pneumonia		
	IS_A bacterial infectious disease	
	IS_A infective pneumonia	
	causative agent bacteria	
	finding site lung structure	
	Fully-defined	All pneumonias with causative agent "bacteria" are "bacterial pneumonia."

[5] Note that the Relationships shown in the table and diagram are not the definitive released Relationships of these Concepts. They have been simplified to illustrate particular points in the text.

Table 5.4: Canonical form applied to an example Concept[6]

Disease		
	IS_A SNOMED CT Concept	
	Primitive	Not all SNOMED CT Concepts are diseases. No defining characteristics are included to specify what makes something a disease.
Infectious disease		
	IS_A disease	
	causative agent infectious agent	
	Primitive	Not all diseases with causative agent “infectious agent” are “bacterial infectious diseases.” For example “rheumatic heart disease” has causative agent “streptococcus” but is not an “infectious disease.
bacterial infectious disease		
	IS_A infectious disease	
	causative agent bacteria	
	Fully-defined	All “infectious diseases” with causative agent “bacteria” are “bacterial infectious diseases”
disease of respiratory system		
	IS_A disease	
	finding site respiratory system structure	
	Fully-defined	All diseases with associated topography “respiratory system structure” are “diseases of respiratory system.”
disease of lung		
	IS_A disease	
	finding site lung structure	
	Fully-defined	All diseases with associated topography “lung” are “diseases of lung.”
Pneumonia		
	IS_A disease	
	finding site lung structure	
	Primitive	Not all diseases of lung structures are “pneumonia.” No additional defining characteristics specify what attributes are needed to specify “pneumonia.”
infective pneumonia		
	IS_A infectious disease	
	IS_A pneumonia	
	Fully-defined	All pneumonias that are also infectious diseases are “infective pneumonia.”
bacterial pneumonia		
	IS_A infectious disease	
	IS_A pneumonia	
	causative agent bacteria	
	Fully-defined	All pneumonias with causative agent “bacteria” are “bacterial pneumonia.”

[6] Note that the Relationships shown in the table and diagram are not the definitive released Relationships of these Concepts. They have been simplified to illustrate particular points in the text.

5.4.5 Qualifiers and refinement

Qualifiers and refinable definitions

A qualifying characteristic is an attribute that may have one of several possible values for a particular Concept. If a particular qualifier is applied to a Concept, the resulting expression represents a more tightly defined subtype of that Concept.

> **Example:**
>
> It might be possible to qualify a disorder such as "*bacterial pneumonia*" according to its clinical course ("acute" or "chronic") or severity ("mild," "moderate" or "severe"). With appropriate qualifiers, "*injury of skin of the left side of face*" could then be represented even if a single ConceptID cannot express this.

A similar tightening of the definition of a Concept can be achieved by allowing one or more of the defining characteristics associated with a Concept to be refined. A defining characteristic is refined by an expression that applies a specified subtype of the value stated in the definition.

> **Example:**
>
> "*Fracture of bone*" could be refined by qualifying it with the FINDING SITE "tibia" to represent the Concept "*Fracture of the tibia*."

5.4.6 Primitive and fully-defined Concepts

A Concept is considered to be fully defined if its defining characteristics are sufficient to define it relative to its immediate supertype(s). A Concept which is not fully defined is Primitive and this is indicated by the value of the IsPrimitive field.

> **Example:**
>
> "*Pneumonia*" is a lung disease but unless defining characteristics are specified that effectively distinguish "*pneumonia*" from other lung diseases then it is regarded as a primitive Concept.

If a Concept is primitive then the defining characteristics for that Concept are incomplete. It is not possible to automatically compute that a Concept represented as a post-coordinated combination of several Concepts is or is not a subtype of a particular primitive Concept.

> **Example:**
>
> The Concept "*lung disease*" qualified by "causative agent" = "bacteria" may be "*pneumonia*" but could also be "*bronchitis*."

In contrast if a Concept is fully defined it is possible to state that any Concept represented as a combination of the same defining characteristics is equivalent to or a subtype of that Concept.

> **Example:**
>
> Assume that the Concept "*bacterial pneumonia*" is fully defined as "*infective pneumonia*" with "causative agent" = "bacteria" and that "*pneumococcus*" is a "bacteria." It then follows that the post coordinated representation of "*pneumococcal pneumonia*" as "*infective pneumonia*" with "causative agent" = "pneumococcus" is computably a subtype of "*bacterial pneumonia*."

5.4.7 Active Concepts

Current Concepts

SNOMED CT Concepts that are intended for active use are referred to as Active Concepts.

Most Active Concepts are suitable for general use and have the ConceptStatus "current."

Other Active Concepts are intended for active use in particular circumstances. These Concepts have the ConceptStatus value "Limited" or "Pending move." The significance of these status values is described in the following sections.

Limited Concepts

Concepts with the ConceptStatus "Limited" are released as Active Concepts only because they exist in classifications or administrative groupings. These may be needed to meet management or epidemiological objectives. However, they are not recommended for general use due to limitations in the consistency of their interpretation.

These Concepts may represent aggregations of dissimilar Concepts or exception categories intended for otherwise unclassifiable information. Their meaning is dependent on context and potentially subject to change or reinterpretation.

> **Example:**
>
> Classifications like ICD10 include terms containing phrases such as "not otherwise specified" or "not elsewhere classified."
>
> - These are only valid in respect of a particular classification. Another classification or another version of the same classification may include a more precisely defined category that narrows the scope of "not otherwise specified."
> - Many Concepts that would fall into "not otherwise classified" categories in ICD10 can be more precisely represented by another SNOMED CT Concept. From a SNOMED CT perspective the phrase "not otherwise classified" is therefore open to misinterpretation.

Concepts with the status "Limited" should not be used in clinical records. However, applications should be able to present these Concepts for use where appropriate to meet particular requirements for direct use of classification categories.

Concepts Pending Movement to other namespaces

Concepts with the ConceptStatus "Pending move" are released as Active Concepts while waiting for a transitional period after which they are expected to be released in another namespace.

SNOMED CT supports the inclusion of Extensions to the terminology managed by different organizations. Concepts released by different organizations have ConceptIDs allocated within different namespaces. In some situations it is appropriate to transfer maintenance of one or more Concepts between separately managed namespaces.

After a transfer the original Concept becomes inactive with the ConceptStatus "Moved elsewhere." However, the process of release and installation of updates from different extensions will typically be asynchronous. Thus a Concept move may be initiated in the original namespace but not yet completed in the target namespace. In this situation, the Concept is released with the ConceptStatus “Pending move.” In addition to their active Descriptions and Relationships, these Concepts have a “Moved to” Relationship indicating the target namespace for the future maintenance of the Concept.

If the implementer or user has access to the replacement Concept in the target namespace then a Concept with this status may be treated as inactivated. However, unless and until such Concept is available, a “Pending move” Concept should be treated as an Active Concept.

5.4.8 Inactive Concepts

Many Concepts released in SNOMED CT are, for one reason or another, not recommended for active use. These Concepts are released as part of SNOMED CT to support legacy data.

The ConceptStatus indicates the reason for inactivation of a Concept. These include the following:

- **Retired**: The Concept has been withdrawn without a specified reason.
- **Duplicate**: The Concept has been withdrawn from active use because it duplicates another Concept.
- **Outdated**: The Concept has been withdrawn from active use because it is no longer recognized as a valid clinical Concept.
- **Ambiguous**: The Concept has been withdrawn from active use because it is inherently ambiguous.
- **Erroneous**: The Concept has been withdrawn from active use because it contains an error. A corrected, but otherwise similar, Concept has been added to replace it.
- **Moved elsewhere**: The Concept is no longer maintained by the organization responsible for the namespace of this ConceptID. The Concept has been moved to another namespace indicated by an associated “Moved to” Relationship.

All Inactive Concepts are subtypes of an appropriate subtype of the Special Concept “Inactive concept.” They do no participate in any other subtype Relationships and have no defining characteristic.

5.4.9 Historical relationships

Roles of historical relationships

Historical relationships link Inactive Concepts to the overall semantic structure of SNOMED CT. These specialized Relationships should be used to map legacy data into the current subtype hierarchy of SNOMED CT:

- An “erroneous” Concept is related to the corrected Concept that replaces it by a “REPLACED BY” Relationship.
- A “duplicate” Concept is related to the Concept that it was found to duplicate by a “SAME AS” Relationship.
- An “ambiguous” Concept is related to the Concept (s) that represent its possible disambiguated meanings by one or more “MAY BE A“ Relationships.
- Any Inactive Concepts may be related to another Inactive Concepts that was previously considered to be its supertype by a “WAS A” Relationship.
    - “WAS A” Relationships are only included where they add value.

Example:

Where a Concept and its previously released subtypes are inactivated due to a similar ambiguity.

- A Concept that has the status "Moved Elsewhere" or "Pending Move" is related to a Namespace Concept by a "MOVED TO" Relationship. This identifies the namespace to which the Concept has been moved or will be moved.
- A Concept that replaces a Concept moved from another namespace has a "MOVED FROM" Relationship referring to the original Concept in its original namespace.
  - If the Extension from which a Concept originated is not installed the target Concept will not be available. This is not an error and can be safely ignored. The information in the Relationship is only of interest to users of the original Concept.
  - If the Concept has been moved to the SNOMED CT core namespace the "MOVED FROM" Relationship should be maintained in the originating Extension.

5.5 Navigation Concepts and navigation links

5.5.1 Role of navigation Concepts and navigation links

The subtype Relationships described in [5.4.2] provide a logical semantic hierarchy. Often it is possible to view parts of the terminology and select particular Concepts by navigating through this subtype hierarchy. However, there are many situations in which the pure subtype hierarchy does not provide an ideal route for navigating the hierarchy.

Navigation links are used to provide an alternative route through parts of the terminology. A navigation link can link any two Concepts together to identify a useful route for navigation. Each of the navigation links is directional, linking a navigational parent Concept to a more refined navigational child Concept. However, unlike the subtype relationship the presence or absence of a navigation link neither adds to nor subtracts from the definition of either of the Concepts that it links.

Navigation Concepts provide additional intermediate nodes in the navigation structure. A navigation Concept plays no part in the semantic definitions of any other Concept.

5.5.2 Representation of navigational Concepts

Navigation Concepts are subtype children of the Special Concept "Navigation concept." They have no subtypes of their own.

The "Navigation concept" has a designated ConceptID that is documented in Appendix B.

5.5.3 Representation of navigational links

Navigation links are represented as rows of a Navigation Subset. Several Navigation Subsets may be released, each representing an alternative hierarchical view of SNOMED CT. Each Subset Member identifies a parent Concept (MemberID), an order (MemberStatus) and a child Concept (LinkedID).

Breaking down a subtype into manageable categories

Some Concepts have a large number of subtype children that cannot be logically divided into intermediate subtypes. At the user interface these result in long lists of options, which are difficult to visualize and navigate. Navigational Concepts with appropriate navigational links to the supertype parent and its subtype children provide an intermediate layer without disrupting the semantic definitions.

The “clinical finding” top-level Concept has a large number of subtype children. Intermediate navigation Concepts group some of these together in a convenient way.

Example:

Three subtypes related to pregnancy are grouped together under a single natural navigational Concept:

- *Disorder of pregnancy / labor / delivery / puerperium [navigation concept]*
- *Disorder of pregnancy*
- *Disorder of labor / delivery*
- *Disorder of puerperium*

Bypassing levels in the subtype hierarchy

Some Concepts that a user may consider to be part of a rational set of choices may be found at different levels in the subtype hierarchy. This may occur before some have intervening subtypes (which are rarely used for data entry). Furthermore, the introduction of a new Concept may cause a change in the levels in the hierarchy as such choices are found in a new release. Navigation links can “bypass” a level in the subtype hierarchy to bring together a rational set of choices without altering the definition of a Concept.

Example:

While it is semantically correct to nest “common cold” in the following subtype hierarchy, a user may reasonably expect to see “common cold” as an immediate navigational child of “*upper respiratory infection*.”

upper respiratory infection
  viral upper respiratory infection
    common cold

Linking related Concepts of different types

Navigational links can also be used to provide access to connected Concepts even when they are from different hierarchy branches.

Example:

A navigation links could associate:

- “hypertension” (the disorder) with “blood pressure” (the observation).
- “cataract” (disorder / finding) with “cataract surgery” (the procedure).

Ordering the display of subtypes

Sibling Concepts in a subtype hierarchy are not ordered. However, at the user interface a particular order may be useful to highlight commonly used Concepts or to mirror a conventional ordering.

Example:

Vertebrae, cranial nerves, disease stages, etc.

Navigational links are ordered and are used to impose order, even when the set of navigational children is the same as the set of subtype children.

Providing alternative hierarchies

The subtype hierarchy is logically defined and there can only be one such hierarchy. However, as navigation hierarchies have no definitional consequences, it is possible to have different hierarchies for different groups of users with differing needs.

Initial releases of SNOMED CT will contain a single set of navigation links but those engaged in technical implementation should be aware that in the future there may be separate sets of navigation links for use in different environments.

6 Terminology Services Guide

6.1 Introduction

6.1.1 Definition

In the context of this document, this section will only address how terminology services should provide access to SNOMED CT terminology resources. Additional terminology services that provide access to other terminologies and classifications may be required to support a more complete implementation.

6.1.2 Significance

An appropriate, reliable and efficient set of terminology services is an essential foundation for any SNOMED CT Enabled Application.

Superficially, terminology services may seem a little different than providing a database table lookup. However, more sophisticated solutions are needed to deal with the size of SNOMED CT (e.g. more than half a million terms) and to make effective use of the power of its semantic structures.

6.1.3 Development options

This part of the guide is intended for those developing or using terminology services either as an integral part of an application or as a separate software component accessed via an Application Programming Interface (API). These options are outlined in [4.5] and discussed further in [6.10].

6.2 Foundation Terminology Services

6.2.1 Introduction

The following list summarizes a set of essential services that any terminology server is likely to require. Some of these are reiterated in more detail in subsequent sections. In other cases, support for a more complex service presumes the existence of one or more of these foundation services.

6.2.2 Access to release information

Terminology servers should enable client applications and users to access the current SNOMED CT release information [5.4.1].

6.2.3 Access to components

Access to concepts

A terminology server should enable client applications to rapidly find any *Concept* by any of the following criteria:

- ConceptID
- SNOMEDID
- CTV3ID

Once a Concept has been found, the client application should be able to read the values. Any of the properties of that Concept included in the Concepts Table should be accessible.

Access to descriptions

A terminology server should enable client applications to rapidly find any Description or set of Descriptions by any of the following criteria:

- DescriptionID
- ConceptID
- ConceptID and DescriptionType

Once a Description has been found the client application should be able to read the values of any of the properties of that Description included in the Descriptions Table.

Access to relationships

A terminology server should enable a client application to rapidly find any Relationship or set of Relationships by any of the following criteria:

- ConceptID1
- ConceptID1, CharacteristicType and RelationshipType
- ConceptID1, CharacteristicType, Relationship group and RelationshipType
- ConceptID2
- ConceptID2, CharacteristicType, and RelationshipType

Once a Relationship has been found the client application should be able to read the values of any of the properties of that Relationship included in the Relationships Table.

6.2.4 Access to essential concept identifiers

Terminology servers should provide access to the ConceptIDs that represent the following Concepts which have structurally significant roles within the terminology.

- The root Concept
- The subtype (IS_A) relationship
- Top-level Concepts with structural roles
  - Linkage concept
  - Qualifier value
  - Special concept
- The Special Concepts:
  - Inactive concept
  - Namespace concept
  - Navigational concept

6.2.5 Testing and traversing subtype relationships

Access to hierarchically related concepts

Terminology servers should enable client applications to access collections of Concepts that are related to a specified Concept as:

- Subtype children
- Subtype descendents (includes all generations of children)
- Supertype parents
- Supertype ancestors (includes all previous generations of parents)

Top-level ancestor checking

Terminology servers should allow client applications to rapidly determine the top-level Concept that is the supertype ancestor for any specified Concept.

Each Concept has only one top-level supertype and this represents the semantic-type of the Concept.

Navigation concept checking

Terminology servers allow client applications to determine whether a specified Concept is a navigation Concept.

Subtype descendant testing

Terminology servers should be able to test whether any specified Concept is a descendant subtype of another specified Concept.

Subtype search scope restriction

Terminology servers should be able to restrict searches so that they return only those Concepts and (or their associated Descriptions) that are subtype descendants of a specified Concept.

Subtype search scope restriction is particularly valuable with respect to top-level Concepts. For example, when searching for a procedure it is useful to be able to exclude disorders or findings that may contain similar words of phrases.

Generalizing subtype search scope restriction to other nodes in the subtype hierarchy may significantly enhance usability in some situations.

> **Example:**
>
> When undertaking an ophthalmologic examination, a search for findings could be constrained to findings related to the eye, increasing the specificity of results of searches for phrases containing the word "fundus."

Subtype test implementation issues

Explicit testing of subtype relationships

It is possible to meet all the requirements in this section by following sequences of "IS_A" Relationships up or down from a Concept until a particular target is reached or until all possible paths have been exhausted. However, some functions require a large number of such tests to be undertaken with an almost immediate response.

Some databases include additional features to support the recursive testing of a chain of hierarchical relationships. Other methods of optimization that may be applied to allow more rapid computation of subtype descendant relationships are outlined in the following subsections.

Semantic type identifiers

The internal representation of each Concept can be extended to include identifiers representing the top-level Concept (or semantic type) from which it is descended.

Any Concepts that do not share the same top-level Concept as the putative ancestor are not descendants and can be excluded from further testing. Those that do share flags can be tested by following the sub-type Relationships.

Typically this will reduce the number of exhaustive tests to be performed by a factor of between ten and twenty.

Hierarchy flags

The internal representation of each Concept can be extended with a set of flags representing a set of higher-level supertype Concepts from which it is descended.

> **Example:**
>
> This might include flags for the top-level Concepts plus the most populous branches from the next couple of levels of the hierarchy.

Any Concepts that do not share the flag settings of the putative ancestor are not descendants and can be excluded from further testing. Those that do share flags can be tested by following the subtype Relationships. A further refinement is that there is no need to continue to follow any pathway that leads through a Concept that no longer shares the flags of the ancestor.

The effect of this depends on the number and specificity of the flags. Experiments have shown that a set of 64 hierarchy flags reduces the average number of exhaustive tests required by a factor of nearly a hundred.

Branch numbering

The internal representation of each Concept can be extended to include a branch-number and a set of branch-number-ranges.

A branch-numbering algorithm can then be applied when each release of SNOMED CT is imported.

A typical branch-numbering algorithm processes the subtype hierarchy in the following way:

- A depth first tree walk is performed starting from the root Concept (branch-number 1) and an incrementing number is applied to each Concept when it is encountered for the first time.
- After the branch numbers have been computed a further tree walk allocates one or more branch-number ranges to each Concept with any subtype descendants.
    - Many Concepts will have a single branch number range containing all their descendants.
    - Some Concepts will have several non-contiguous ranges of descendant Concept branch numbers.
        - This is because a Concept may have multiple supertypes. Therefore, the descendants of a Concept may have branch numbers that were allocated as a result of their relationship to another ancestor Concept. However, the path from any Concept to the root Concept always converges at or before the top-level Concept. Therefore, multiple ranges coalesce when reaching more general common supertype ancestors.
- At run time, rather than needing to traverse many subtype Relationships, the branch number of each Concept is tested for inclusion in the branch number range of the putative ancestor.

This approach removes the need for exhaustive testing of subtype Relationships. The disadvantages are a relatively complex build process that must be repeated for each release or update and a requirement for the internal Concept representation to accommodate a variable length representation of branch number ranges.

6.3 Representation of the SNOMED CT Resource

6.3.1 Introduction

One of the first steps in the design and development of a terminology server is to decide how the SNOMED CT resources will be stored and accessed. Options include use of a relational database, an object database or a proprietary file structure. The decision will be influenced by the experience of the developers and the technical environment in which development is undertaken. The following sections briefly outline some of the options and identify general issues that should be considered when making this decision.

Those intending to use a third-party terminology server are unlikely to be faced directly with the decision of how to represent the information. However, they will be interested in the consequences of these decisions in terms of performance and the overall size of the installed SNOMED CT resource. If a terminology server uses a particular relational database with an additional licensing or technical support cost this may also affect its suitability.

6.3.2 Distribution files

SNOMED CT is provided as a set of distribution files. These are specified in detail in the referenced documents:

- Required for all SNOMED CT terminology servers
  - Core Structure
  - Subset Mechanisms
- Required to support mapping to classifications and other coding schemes
  - Cross Mapping Tables
- Required to support version control services
  - History Tables
- Added value files (the data contained is derived from other released files). These are SNOMED CT Derivatives.
  - Word and Phrase Search Tables (part of Developer Toolkit)
  - Canonical Table

All of these materials are distributed as files that:

- Contain data tables presented in a tab-delimited form[7].
- Are represented in accordance with the Unicode UTF-8 specification.

6.3.3 Direct use of distributed files in a relational database

It is possible to import the distributed files directly into a database schema and to use this as the SNOMED CT resource at the heart of a terminology server.

Appendix A contains a summary of all the distribution tables with advice on additional keys and indices that may be required or useful to allow effective use of SNOMED CT in this basic relational form.

This direct use of distributed files in a relational database has the advantage of allowing simple installation. However, it may not be the most efficient approach in terms of performance or file

[7] Subsets are available as tab-delimited representations (Subsets Tables) and in an enhanced XML representation (Subset Definition Files).

size. Some terminology services require relatively complex queries with multiple joins, and need to be completed in fractions of a second to provide an acceptable user interface.

Example:

- Find all the Descriptions associated with a particular Concept.
- Find all the Concepts that are subtypes of a particular Concept.
- Search for an Active Concept in a specified Subset that matches a particular word order independent text phrase.
- Establish whether two Concepts are equivalent according to their defining characteristics.

The performance criteria of searches and joins in very large relational databases vary significantly. Therefore, different optimizations may need to be used to achieve acceptable response times according to the nature of the relational database system.

6.3.4 Alternative relational structure

There is no requirement to use the data structure as distributed. Other structures can be used provided they are able to deliver the range of terminology services required. Options include:

- More (or less) normalized representations of the model.

Examples:

Excluding the Fully Specified Name field from the Concepts table.

- Increases normalization of the data since the Fully Specified Name is also available in the Description table.

Replicating the "IS_A" Relationships from the Relationships table with the source and target ConceptID fields reversed.

- Decreases normalization of the data but might have some performance benefit.

- Omission of some of the tables.

Example:

If a terminology server is not going to support cross mapping, the Cross Map Set and Cross Map Target tables are not required.

- Replacement of some of the supporting tables with proprietary alternatives

Example:

The word search support tables could be completely replaced by other tables or indices generated by the terminology server when loading the distribution files.

6.3.5 Use of non-relational structures

Although the primary distribution format is relational, this does not require terminology servers to utilize a relational database as the primary or only storage format. Review of the requirements for terminology services may suggest other structures such as object-oriented databases, use of Extensible Mark-up Language (XML) or proprietary structures. These structures may be employed separately or, in some cases, in combination with a relational database.

A possible conclusion from analysis is that most terminology services are concept-centered. One structure that reflects this is a hierarchy in which all the information about each Concept is kept together as a discrete unit.

Example:

SNOMED CT

- [1..n] Concepts
- [1..n] Descriptions
- [1..n] Relationships
- [1..n] Cross Maps

Such a structure needs support from indices and/or other data access tools to allow searches. However, it has the merit that whenever a Concept is accessed, all the relevant information is immediately at hand without following relational joins to multiple rows in other tables.

6.4 Importing and Updating SNOMED Clinical Terms

6.4.1 Importing distribution files

Requirements for importing distribution files

Terminology servers should be able to import SNOMED CT distribution files into the internal form used for the SNOMED CT resource.

The import process should be automated with user intervention limited to selection of *Subsets* and/or configuration options.

The import process should import:

- The core files (Concepts, Descriptions, and Relationships)
- A selected set of Subsets
  - This import may be either from
    - A selected set of Subset Definition Files
    - Selected rows in the Subsets Table and Subset Members Table
  - Some Subsets may be used to filter the import of other distribution files[8].
    - Descriptions may be filtered to limit coverage to a specified language or dialect.
    - Concepts may be filtered to limit coverage to Concepts relevant to human medicine by excluding those only relevant to veterinary medicine.
    - Relationships associated with any Concept that are excluded by Subset filtering should also be filtered out.
  - Language Subset may alter the DescriptionType and/or LanguageCode values in the distributed Descriptions Table.
- Any other distribution files needed to support the required set of services.

[8] See also guidelines on implementation of Subsets. Language Subsets can be used in single language implementation by filtering and applying the DescriptionType and LanguageCodes during the import process. However, multi-lingual installations may require dynamic application of the alternative values applicable to the supported languages and dialects.

Checking during the import process

The import process should check the imported data to confirm that:

- The distribution files imported all parts of the same release.
- All SCTIDs have:
  - A partition-identifier appropriate to the field.
  - A valid check-digit.

Other consistency checks may also be applied to ensure the integrity of the data.

Pre-processing of distribution files by terminology server suppliers

The import process may be time-consuming due to the need to build indices or other data structures. It may also require substantial spare storage capacity for temporary files. Therefore a terminology server provider may choose to pre-import the distribution files and provide them to users in pre-prepared form. However, an import facility should also be available in a suitably secured form to end-user organizations, to enable installation and maintenance of Extensions.

6.4.2 Importing new releases and updates

Introduction

This section is concerned with updates to the SNOMED CT resource accessed by the terminology server. A separate section addresses issues related to management of SNOMED CT encoded records, queries and protocols following a new SNOMED CT release (see [7.3.2]).

Update release formats

Each release of the core content of SNOMED CT will be distributed in full rather than as a set of changes and additions. However, history information is provided. The structure of the history files is described in detail in the Technical Reference Guide.

Future options may include distribution of change set files that include changes between versions, provided in addition to the full release. This distribution form may also be appropriate for inter-version additions to cover newly requested Concepts. At this stage there are no detailed specifications for these types of updates. However, a Concept centered XML representation for such change sets is under discussion.

Updating from distribution files

Updating a terminology server from a new release of distribution files is similar to the original import. However, the import process should support the retention and/or reinstatement of local configurations such as installed Extensions and selected Subsets.

An alternative to repeating the full import process is to read and apply the changes indicated in the Component History Table. The distribution files are still required as the source for additional components that have been added between releases. Also, the Component History Table only indicates changes to Concepts and Descriptions. However, changes to the distribution format to allow better representation of all SNOMED CT component changes is under discussion.

This process is outlined as follows:

- Read the Component History Table
  - Identify all changes since the last release used to update this server.
- For each of these changes check the ChangeType value.

- If the ChangeType is 1 (status change), apply the new status to the component in the terminology server resource.
    - For example, if the new status of a Concept is Duplicate, this status is applied to the identified Concept.
    - For many inactivated concepts, the Relationships file can be used to identify an inactive concept that may be a candidate to use in place of the inactivated concept.
- If the ChangeType is 0 (added), find the relevant component in the distribution files and add this to the terminology server resource.
- If the ChangeType is 2 (minor change), find the relevant component in the distribution files and apply the new data to the component in the terminology server resource.

This approach is more complex than importing the new release. However, it may make it easier to retain local configuration data (e.g. selected Subsets and any Extensions previously installed).

If a terminology server suppliers pre-imports SNOMED CT releases for distribution to their users, this approach facilitates the production of re-distributable updates.

Changes to relationships

Unlike other core files the released Relationships Table has no status field and only active Relationships are distributed. Therefore Relationship changes are not recorded in the Component History Table and requires a full import of this table or the need for an intermediate table to determine which Relationships were added and removed in the distribution.

6.4.3 Importing Extensions

Terminology servers should be able to import Extensions.

The process of importing an Extension is similar to importing the main distribution files. However, some additional functionality is required to ensure appropriate installation, maintenance and use of Extensions. Applications should:

- Allow the users or user communities to specify the Extensions to be recognized by their systems. Before recognizing any Extension, users should check that:
    - The Extension has been supplied by the IHTSDO or another organization authorized by the IHTSDO to provide such Extensions.
    - You are satisfied with the quality control procedures of the providing organization.
        - Authorization of an organization to produce Extensions does not imply any seal of approval related to the quality of Extensions provided by those organizations.
        - Installation of Extensions is done entirely at the risk of the user subject to their license agreement with the provider of the Extension and/or the application developer.
- Check each Extension prior to installation to ensure that:
    - It is one of the Extensions recognized by the user.
    - Any dependencies of the Extension have been met. These dependencies may include:
        - Installation of a particular SNOMED CT release;
        - Prior installation of other Extensions.

  - The installation procedure has pre-checked all Components in the Extension to ensure that:
    - ° All ComponentIDs
      - Are unique
      - Have partition-identifier and namespace-identifier values appropriate to the provider of that Extension
      - Have a valid check-digit
    - ° All fields meet data type, size and value constraints specified for the relevant tables.

Caution:

If any Components fail any of these tests the entire Extension must be rejected.

- Import Extension files in a way that replaces earlier versions of the same Extension but add to, rather than replacing, the SNOMED CT core distribution and other installed Extensions.
- Reject, highlight or apply other agreed business rules to information received by the system that contains SCTIDs for Components from namespaces that are not in the list, or recognized Extensions.

6.4.4 Providing access to history information

When a new release is installed, an application providing record services may need to update records, indices, queries, or protocols to take account of changes to the terminology.

To support these maintenance activities a terminology server should provide access to:

- Information in the Component History Table indicating which components have been changed since the previous release.
- Information in the References Table identifying components which replace or were duplicated by inactive Components.
  - Not applicable to Concepts (see next bullet below)
- Information in the Relationships Table identifying Concepts that:
  - Replace an “erroneous” Concept
    - ° RelationshipType “REPLACED BY”
  - Are identical to a “duplicate” Concept
    - ° RelationshipType “SAME AS”
  - Are potential meanings of an “ambiguous” Concept
    - ° RelationshipType “MAY BE A”
  - Were previously subtypes of an Inactive Concept
    - ° RelationshipType “WAS A” (Future Use)

Access to the Component History Table and References Table may only be required to support for a time a limited set of maintenance activities following each release. However, access to information about Active Concepts related to Inactive Concepts is often required for day-to-day operation of the application. For example, it will affect the selective retrieval and analysis of data information originally recorded with a Concept that is now inactive.

6.5 Text Searches

6.5.1 Introduction

Effective implementation of SNOMED CT depends on the speed and simplicity with which users can locate the terms and concepts that they wish to use. A busy clinical user may become frustrated if the content they need cannot be quickly located when they search using familiar words or phrases. For this reason an efficient search strategy should address the following issues:

- Speed of search
  - Search speed should be optimized by use of appropriate indexes
- Search should not be too sensitive to word order or exact phrasing
  - Search should be insensitive to word-order variants
    - For example, “head pain” for “pain in the head”
  - Allow use of acronyms or abbreviations for frequently used terms
    - For example, “MI” for “myocardial infarction” or “mitral incompetence”
  - Search should take account of word form variants
    - For example, “inflamed”, “inflammatory”, “inflammation”
- Excessive search results should not hinder selection of the required concept
  - When several synonyms of the same concept match the search key, only one should be displayed.

The purpose of this section of the implementation guide is to describe strategies a developer might use to implement the search requirements outlined above.

The SNOMED CT Developer Toolkit contains several files, which help to support efficient search mechanisms. These include the Excluded Words Table, four keyword indexes and the Word Equivalents Table summarized by Table A.4 in Appendix A.

6.5.2 Single keyword index

The single keyword table, (DescWordKey), provides a pointer from each keyword used in any Description, to the Descriptions in which that keyword is used. The purpose of the single keyword index is to support a search capability, which is independent of the order in which words appear in a description. The single keyword index represents the minimum necessary supporting structure for searches on SNOMED CT content. Searches involving target words that appear in many descriptions may be unacceptably slow if searches are carried out using the single keyword index alone. Developers wishing to produce applications with faster search times are encouraged to supplement their system with the multiple keyword index (DescDualkey) described in [6.5.3].

Note that some words that are used in description are linking words, which are unlikely to be in the target of a search. These words are not considered to be keywords and may be excluded from the keyword index. They are found in Excluded Words File.

Generating the single keyword index

Although single keyword indexes are available in the Developer Toolkit for all core English language descriptions, developers will need to know how to add keyword entries for any locally generated descriptions added as part of an Extension. An extended toolkit is now offered that contains the same programs used to generate the core keyword indexes.

Entries may be added to the single keyword table by following the method outlined below.

For each description, parse the text of the term:

- To avoid inappropriate case mismatches, convert all characters to the same case.
- Extract words by breaking at spaces, punctuation marks, and brackets.
- For each word:
  - If the word is not in a list of excluded words, add a row to keyword table.

Example: Generation of keywords for a sample description

DescriptionID	ConceptID	Term
22565018	13185000	Pyrogallol 1,2-oxygenase

- Convert all characters to the same case
  “Pyrogallol 1,2-oxygenase” ➔ “PYROGALLOL 1,2-OXYGENASE”
- Extract words by breaking at spaces, punctuation marks, and brackets.
  “PYROGALLOL 1,2-OXYGENASE” ➔ (1) = “PYROGALLOL”
  (2) = “1”
  (3) = “2”
  (4) = “OXYGENASE”
- For each word:
  - If the word is not in a list of excluded words, and length of word > 1, and first character is not numeric:
    - Add a row to keyword tables.
    - Only the first eight characters are used in the keyword.

Table DescKey Words

KeyWord	DescriptionID
PYROGALL	2265018
OXYGENAS	22565018

Table ConcKeyWords

KeyWord	ConceptID
PYROGALL	13185000
OXYGENAS	13185000

Search using the single keyword index

A single keyword search may be conducted as follows:

- The user-typed search string is converted to consistent case.
- The string is parsed, breaking at spaces and punctuation characters.
- One word is selected from the parsed word list to use as a look-up on the single keyword index.
- Look-up on the single keyword index may be “exact” or “starts with,” depending on wild card conventions used in the search string.

Example: Search using single key-word index

The user searches for “Hip* replacement*” (where “*” represents the wild card for any number of extra characters).

- The user-typed search string is converted to consistent case.
  “Hip* replacement” ➔ “HIP* REPLACEMENT*”

- The string is parsed, breaking at spaces and punctuation characters.
  "HIP* REPLACEMENT*" ➔ (1) "HIP*"
  (2) "REPLACEMENT*"
- Look up "HIP" on the single keyword index using "starts with" query.

Search results

Count	DescriptionID	ConceptID	Term
1	49926016	29836001	Hip
2	196344018	24136001	Hip
3	48043014	736004	Abscess of Hip
4	22723010	13279003	Partial hip replacement by prosthesis
. . . . .	. . . .	. . . .	. . . .
315	160880013	97741000	Hiprex Tablets

Descriptions in the search results are converted to consistent case and screened, to see if they contain any words starting with "REPLACEMENT." Any that do not are not included in the final search results.

Final Search Results

Count	DescriptionID	ConceptID	Term
1	22723010	13279003	Partial hip replacement by prosthesis
2	33592011	19954002	Total replacement of hip with use of methyl methacrylate
3	50150016	29969002	Replacement of acetabulum of hip
4	54398014	32581000	Partial hip replacement by cup with acetabuloplasty
. . . . .	. . . .	. . . .	. . . .
11	183737015	112728000	Total revision of hip replacement with use of methyl methacrylate

6.5.3 Multiple keywords

Introduction

The performance of single keyword searches is highly dependent on the number of candidate descriptions returned by the keyword for subsequent filtering. The extremely high number of matches for some words in common use makes it likely that some searches will be unacceptably slow.

One way to alleviate this problem would be to create a table containing a row for all combinations of word pairs in each description. In some database environments that support optimization of multiple key searches, this may offer no benefits. However, in other environments, such a table may substantially speed searches.

A comprehensive word pair table would be very large. Such a table covering the full content of SNOMED CT would contain approximately 1.5 million unique word pairs and 6 million rows. Limiting the unique keys to the first three letter of each word reduces the table size to a more readily optimized set of keys. This requires the final part of the search to be conducted using text comparison (since the keys are incomplete).

Generating the word pair index

Although Dualkey indexes are available as part of the Developer Toolkit, it is important to know how this table is generated. SNOMED CT users that generate Extensions should follow the method outlined below to generate new entries in the Dualkey index, based on the descriptions in the Extension.

For each description, parse the text of the term:

- To avoid inappropriate case mismatches, convert all characters to the same case.
- Extract words by breaking at spaces, punctuation marks, and brackets.
- For each word of three characters or more that is not in the list of excluded words, extract the first 3 characters, and arrange the word fragments in alphabetical order.
- Generate the dual keys for this description by concatenating each word fragment with those that come after it in the list.
- For each dual key, add a row to the word pair tables.

Example: Generation of keywords for a sample description

DescriptionID	ConceptID	Term
33592011	19954002	Total replacement of hip with use of methyl methacrylate

- To avoid inappropriate case mismatches, convert all characters to the same case: "TOTAL REPLACEMENT OF HIP WITH USE OF METHYLE METHACRYLATE"
- Extract words by breaking at spaces, punctuation marks, and brackets:
  (1) TOTAL
  (2) REPLACEMENT
  (3) OF
  (4) HIP
  (5) WITH
  (6) USE
  (7) OF
  (8) METHYLE
  (9) METHACRYLATE
- For each word of three characters or more, that is not in the list of excluded words, extract the first 3 characters, and arrange the word fragments in alphabetical order:
  (1) HIP
  (2) MET
  (3) REP
  (4) TOT
  (5) USE

Note:

"OF" is less than 3 characters and is an excluded word, "WITH" is an excluded word and "MET" is duplicated, so we only include it once.

- Generate the dual keys for this description by concatenating each word fragment with those that come after it in the list.
- For each dual key, add rows to the word pair tables.

DescDualKey

Dual key	DescriptionID
HIPMET	33592011
HIPREP	33592011
HIPTOT	33592011
HIPUSE	33592011
METREP	33592011
METTOT	33592011
METUSE	33592011
REPTOT	33592011
REPUSE	33592011
TOTUSE	33592011

ConcDualKey

Dual key	ConceptID
HIPMET	19954002
HIPREP	19954002
HIPTOT	19954002
HIPUSE	19954002
METREP	19954002
METTOT	19954002
METUSE	19954002
REPTOT	19954002
REPUSE	19954002
TOTUSE	19954002

Searching for descriptions using the word pair index

A search on the dual key index can only be carried out if the user enters a search string that contains at least two word fragments both of which are three characters or more in length. If the search string does not meet this criterion, the single keyword search mechanism must be used.

- The user-typed search string is converted to consistent case.
- The string is parsed, breaking at spaces and punctuation characters.
- For each word of three characters or more, extract the first 3 characters, and arrange the word fragments in alphabetical order.
- Create a dual key by concatenating the first two 3 letter word fragments.
- Use this dual key to look up exact matches on the word pair index.
- Descriptions found by searching on the word pair index are screened, to see if they contain the complete words in the original search string

Example: Search using word pair index

User searches for "PYRO* 1 OXYGEN*".

- The string is parsed, breaking at spaces and punctuation characters:
  (1) "PYRO*"
  (2) 1
  (3) "OXY*"
- For each word of three characters or more, extract the first 3 characters, and arrange the word fragments in alphabetical order:

(1) “OXY”

(2) “PYR”

- Create a dual key by concatenating the first two 3 letter word fragments: OXYPYR
- Use this dual key to look up exact matches on the word pair index:

Dual key	DescriptionID	Description
OXYPYR	1969019	o-Pyrocatechuate oxygenase
OXYPYR	22565018	Pyrogallol 1,2-oxygenase
OXYPRY	104951019	2,5-Dihydroxy pyridene oxgenase

- Descriptions found by searching on the word pair index are screened, to see if they contain the complete words in the original search string
  - Description 1969019 is eliminated since it does not contain the word “1”
  - Description 104951019 is eliminated, it does not contain the word “1” or any word beginning with the string “pyro”

6.5.4 Using word equivalents to enhance searches

In healthcare, there are many words with equivalent meanings. Synonyms provide alternative phrases referring to the concept. However, synonyms are not created automatically for every possible combination of words with an equivalent meaning. The success of simple searches using one or more keywords depends on the text of the available descriptions. Therefore searches will fail or will be incomplete where a different equivalent word is used in the search.

For example: “Kidney stone” and “Renal calculus” are synonymous descriptions in SNOMED CT. A search of SNOMED CT for the target phrase “kidney stone fragmentation” yields the result “Percutaneous nephrostomy with fragmentation of kidney stone,” while a search for “Renal stone fragmentation” yields no results.

One way of addressing this problem is to maintain a table of word equivalents. A table of this type is a prerequisite for exhaustive synonym generation. An initial set of word equivalents is included in the SNOMED CT Developer Toolkit. Individual implementers will wish to add additional word equivalents to meet the requirements of their particular medical specialty or user needs. This table is an additional resource to assist searching and parsing of phrases. It need not be a comprehensive dictionary of words. Many searches can be completed without reference to this table so it need not contain every word or equivalent phrase used in SNOMED CT.

Several factors complicate the initial population and subsequent use of the word equivalents table:

- A phrase of two or more words may be equivalent to a single word.

Example:

“Endoscopic esophagus examination” is equivalent to “esophagoscopy”

- A word may have more than one meaning, and in this, only one meaning of a pair of words may be equivalent. Thus an apparent enhancement of a search may in practice lose some of the specificity of the intended search.

Example:

"Tap" and "aspiration" are equivalent in the context of terms such as "pleural tap", "pleural aspiration", but not in the context of a "patella tap", a physical "tap" on a bag or catheter, or the clinical disorder "neonatal aspiration syndrome."

- When searching using incomplete words and/or wildcards, use of word equivalents may impede effective searches by increasing the number of spurious potential matches. This either extends the processing required to filter the real matches from the potential matches or increases the length of the list of choices presented to the user.

A wise system developer will allow the user to customize their search options, enabling searches to be narrowed, or extended to meet the needs of varying circumstances.

Example: Using word equivalents table to extend a failed search

A system user enters the search string "Fragmentation of renal calculus;" the search returns no results. The search application that the user has been provided with has the option to extend the search by using the word equivalents table. The user selects this option and searches again using the same search string.

The word equivalents table contains the following relevant entries:

WordBlockNumber	WordText	WordType
1021	KIDNEY	2 (word equivalent)
1021	RENAL	2 (word equivalent)
4430	CALCULUS	2 (word equivalent)
4430	CALCULI	1 (word form variant)
4430	STONE	2 (word equivalent)
9870	RENAL STONE	4 (equivalent phrase)
9870	KIDNEY STONE	4 (equivalent phrase)
9870	KIDNEY CALCULUS	4 (equivalent phrase)
9870	RENAL CALCULUS	4 (equivalent phrase)
9870	NEPHROLITH	2 (word equivalent)

The table is used to make substitutions in the search string to produce all possible unique search variants:

"Fragmentation of renal calculus"

"Fragmentation of renal stone"

"Fragmentation of kidney stone"

"Fragmentation of kidney calculus"

"Fragmentation of Nephrolith"

"Fragmentation of renal calculus"

"Fragmentation of renal calculi"

"Fragmentation of kidney calculi"

These 8 search strings are used as the target phrase for keyword searches on the word pair index. Results from all 8 searches are combined, and duplicate concepts are eliminated, giving the final list of search results.

6.5.5 Rationalizing searches that return duplicate hits

Introduction

In the previous sections of this document, we have considered methods of ensuring that searches on a target phrase maximize the possibility of finding the concept that the system user requires. It is equally important to prevent the search results from containing excessive matches, since these will require filtering by the user, imposing an additional burden. Some strategies for limiting the number of search results displayed are discussed below.

Avoiding multiple hits on the same concept

In many instances several synonyms associated with the same concept contain the same keyword. The designer of search software may consider filtering the output of search results so that only the first matching description for a concept is displayed.

> **Example:**
>
> "Endoscopic examination of the stomach" and "endoscopy of the stomach" are synonyms of the same concept. A search for the target phrase "endo* stomach" would return the first phrase found during the search. The second would be excluded, since it has the same concept identifier as an existing match for this search.

Constraining and extending search parameters

User configurable options may be one way of limiting search results. Three possible methods of limiting search results through user configurable options are suggested here:

- Limiting searches to exact matches unless wild cards are used. A search on a single word may produce many matches if it is assumed that the user is searching for any phrase that contains the target word. Forcing the use of wild cards for this kind of search can help avoid this problem.
- Make searches that include use of "word equivalents" a user configurable option that can be used to extend or constrain a search, as described in the example given in section [6.5.4].
- Display search results a few at a time, with most frequently used descriptions listed first. This option will require the application to track the frequency of term selection so that search results can be sorted in this way.

6.6 Hierarchical Navigation

The most viable aspect of SNOMED CT is its vast number of clinical terms. SNOMED CT is more than just a list of terms however; one of its major strengths lies in the relationships that connect the medical concepts underlying the terminology. This section of the implementation guide describes how these relationships may be used to facilitate the search for the correct term or concept for use in a given context.

For circumstances that do not require the full range of SNOMED CT content, subsets can be used to narrow the displayed content to that required for a given application. This section of the guide describes the use of navigation only subsets to provide application specific customized display hierarchies.

6.6.1 Using "IS_A" relationships for hierarchy navigation

What is the SNOMED hierarchy?

The "SNOMED hierarchy" refers to the organization of concepts in SNOMED CT from the general, at the top of the hierarchy, to the more specific or "granular" at the bottom. The concepts that make up the very top level of the hierarchy are shown in Table 6.1. All other SNOMED CT concepts fall under one or more of these categories.

Table 6.1: The top-level hierarchy

Body structure
Clinical Finding
Environments and geographic location
Event
Linkage concept
Observable entity
Organism
Pharmaceutical / biologic product
Physical force
Physical object
Procedure
Qualifier value
Record artifact
Situation with explicit context
Social context
Specimen
Staging and scales
Substance
Special concept

Several levels of increasingly fine categorization may exist between the top level of the hierarchy and concepts that have sufficient detail to be recorded in a patient's medical record. Figure 6.1 shows the levels of hierarchy that exist between the top-level Concept "Finding/disorder" and the finding "Catatonic reaction." On average SNOMED CT contains 11 levels of hierarchy between a given Concept and the top level.

- SNOMED CT Concept
  - clinical finding
    - clinical history and observation findings
      - functional finding
        - psychological finding
          - catatonic reaction

Figure 6.1: Hierarchy example: Catatonic Reaction

How is hierarchical information represented in the Relationship Table?

The SNOMED CT Relationship table represents relationships between one SNOMED CT concept and another by including a row in the table for each such relationship. The columns “ConceptID1,” “Relationship Type” and “ConceptID2” define the source of the relationship, the kind of relationship that exists and the target of the relationship respectively. Each of these fields, including RelationshipType, is a SNOMED CT Concept identifier. Hierarchical relationships are expressed by linking the source concept to its “parents” i.e. the concept or concepts immediately above it in the hierarchy. The RelationshipType used to make this hierarchical link is known as the “IS_A” relationship.

For example, we can say “Catatonic reaction” “IS_A” “Psychological finding,” this would be expressed in the Relationship Table as follows:

ConceptID1	Relationship Type	ConceptID2
102909009	116680003	116367006

Where:

102909009 is the Concept identifier for “Catatonic reaction”

11668003 is the Concept identifier for the “IS_A” relationship

116367006 is the Concept identifier for “Psychological finding”

Conversely, by inverting the “IS_A” relationship we can find the children of the target Concept, i.e. the Concept or Concepts immediately below it in the hierarchy.

Using “IS_A” relationships to enhance search capabilities

A major purpose of the SNOMED CT hierarchy is to allow aggregation of data for the purposes of reporting and research. These issues are discussed in detail in sections [5.4.2] and [7.4]. This section of the document is concerned with the ways in which the hierarchy can be used to help a SNOMED CT user when they are searching or browsing the terminology.

It is possible to start at the top of hierarchy and navigate from parent to child in order to find a Concept or term in SNOMED. A more efficient approach, however, is to use the hierarchy to supplement a keyword search by enabling the user to look at related Concepts in order to consider them as alternative matches, or to check the context of a search result. The examples below illustrate these two uses of the SNOMED CT hierarchy.

Example: Using the hierarchy to check context

A user wishes to find a description that relates to the condition of a patient who is hypersensitive to an allergen. The user performs a search on the keyword “Hypersensitivity” and finds an exact match. Before the user selects the description for inclusion in the patient record, they check the fully specified name, which is “Sensitivity (finding).” The user then checks the hierarchy and discovers that the selected Concept has “Psychological finding” as an ancestor, which indicates that this is not the correct description to use in this context.

Example: Using the hierarchy to check related Concepts

A user wishes to find a description that relates to the condition of a patient who is hypersensitive to an allergen, as in the example above. This time, a different approach is taken. The user searches for the keyword “allergy,” and finds one Concept having a description that is an exact match. The user then looks at the children of the Concept, i.e. those concepts immediately below it in the hierarchy. One of the children has the preferred description “Contact

Hypersensitivity" which matches the user's intended meaning. The user selects this Concept for inclusion in the patient record.

Using "IS_A" relationships to display hierarchical information in applications

Most visual application development tools contain a component designed to display hierarchical information. The Windows "tree view" component is a good example. Many people will be familiar with this view of information, since it is used in the Microsoft Windows Explorer to display folders and their sub-folders. This kind of visual component is ideal for displaying SNOMED hierarchical relationships (see Figure 6.2).

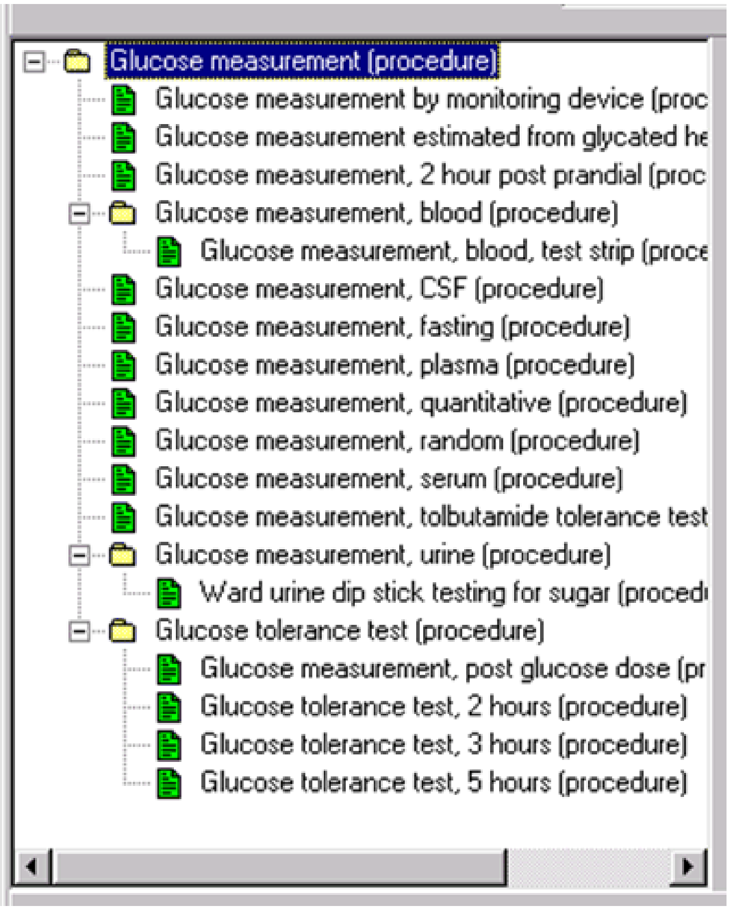


Figure 6.2: SNOMED CT hierarchy represented in a Windows tree view

For those applications that do use a Windows style visual interface, levels of indentation can be used to represent the hierarchy, as shown in Figure 6.3.

```
1    Glucose measurement
2        Glucose measurement by monitoring device
3        Glucose measurement estimated from glycated hemoglobin
4        Glucose measurement, 2 hour post prandial
5        Glucose measurement, blood
6            Glucose measurement, blood, test strip
7        Glucose measurement, CSF
8        Glucose measurement, fasting
9        Glucose measurement, plasma
10       Glucose measurement, quantitative
11       Glucose measurement, random
12       Glucose measurement, serum
13       Glucose measurement, tolbutamide tolerance test
14       Glucose measurement, urine
15           Ward urine dip stick testing for sugar
16       Glucose tolerance test
17           Glucose tolerance test, 3 hours
18           Glucose tolerance test, 5 hours
19           Glucose tolerance test, 2 hours
20           Glucose measurement, post glucose dose
```

Figure 6.3: SNOMED CT hierarchy – plain text with indentation

The process of creating a tree view from the SNOMED CT Relationship table is straightforward as long as a few simple ideas are mastered:

- Since a node cannot be added to the tree until its parent has been added, population of the tree view must start from the top of the hierarchy. This means that when viewing the hierarchy from the bottom up, the tree must be compiled in temporary form from the bottom up, then added to the tree view from the top down.
- Since the depth of the hierarchy is not known in any particular case, operations that iterate up or down the depth of the hierarchy must be done using a recursive algorithm. More detail on recursion in relation to the SNOMED CT hierarchy can be found in Appendix C.
- Concepts with multiple parents in the SNOMED CT hierarchy will appear as multiple nodes in the tree view display.

6.6.2 Using "Part of" relationships for hierarchy navigation

In the proceeding sections we have discussed how the "IS_A" relationship can be used in applications to display the hierarchical information that is part of SNOMED CT. Historically, SNOMED RT used a different relationship, the "Part of" relationship, to represent the hierarchical relationship between a body structure and its parts. For example:

"Right Ventricle" (is) "Part of" "Heart"

These "Part of" relationships are retained for the purposes of backward compatibility with SNOMED RT. They are not defining relationships (as of the January 2005 Release).

Note:

The "Part of" relationship is not currently maintained. It is provided here only as an additional relationship (CharacteristicType=4). Navigation of SNOMED CT for body structures and their parts should be accomplished through the IS_A relationships between structures, entire structures and parts, as explained below.

In SNOMED CT, an entire organ and all its parts are connected via the "IS_A" relationship to a common anatomy structure. For example:

"Right Ventricle" "IS_A" "Heart Structure"

The "IS_A" relationship is the **only** supported representation of the anatomy hierarchy.

6.6.3 Using other relationships to navigate SNOMED CT content

Many SNOMED CT Concepts have relationships with content in other areas of terminology. These Relationships are one of the ways in which SNOMED CT provides computer readable definitions for medical concepts. For example, diseases in SNOMED CT generally have a Relationship to the body site affected by the disorder and a Relationship to the morphology associated with the disease. Procedures in SNOMED might have Relationships to the concept, which defines the type of surgical action being carried and the procedure site, for example. Examples of Relationships for a disease and a procedure are shown below. A full list of the Relationships that can be used for each type of Concept can be found in Table B.5 in Appendix B.

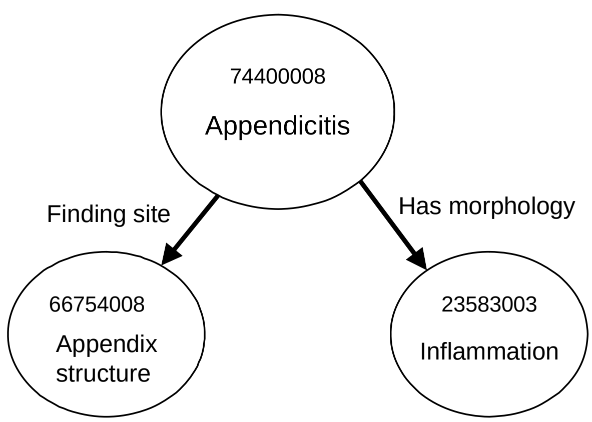


Figure 6.4: Relationship for disease appendicitis

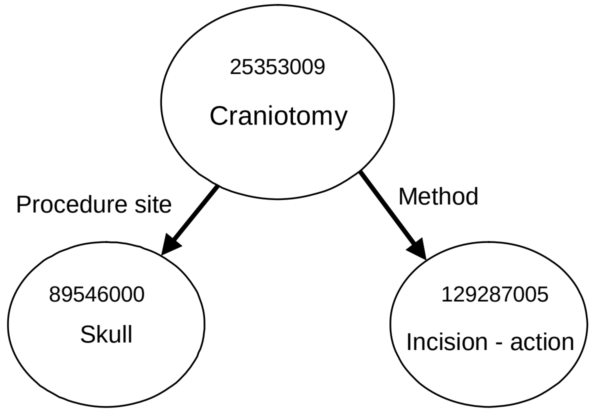


Figure 6.5: Relationships for procedure craniotomy

These Relationships are very useful in the context of data retrieval and analysis, as described in the section [5.4.2] and [7.4]. The Relationships can also be used to aid in the search for specific SNOMED CT Concepts in cases where the term alone may not sufficiently distinguish between choices. For example, a search for all inflammatory diseases of the lung could be carried out as follows:

- Use the hierarchy to compile a list of all Concepts that are lung structures.
- Search for any Concept that has a row in the Relationships table with a Relationship Type of “Disorder site,” and with ConceptID2 included in the list of lung structures.
- Now exclude any procedures from the list that do have the “Associated morphology” “Inflammation” in the Relationship table.
- Final product is a list of all lung disorders that involve inflammation.

To achieve these same results with a string search we would have to perform separate searches for “pneumonia,” “bronchitis,” “pleurisy” and many other conditions that cannot be linked via a sample string search.

6.6.4 Using navigation Subsets to display application-specific hierarchies

The file structure of Subsets is described in detail in the Technical Reference Guide. This section explains how navigation Subsets may be used to create and display a customized hierarchy for use in an application.

Creating a navigation subset

Let us take the example of a hospital specializing in the diagnosis and treatment of cancer. For the purposes of the application, the designers wish to have two top-level categories: “Diagnosis,” including only diagnoses relating to the specialist area; and “Interventions,” including subcategories for chemotherapy, radiation, and surgical treatments.

First we would find the Concepts that we wish to go into the navigation hierarchy. A Subset may only include existing Concepts, by definition, so if any of the Concepts that we require do not exist in the SNOMED CT core files, they will have to be added, either by requesting new concepts from the IHTSDO, or via an approved Extension. For the sake of our example, let us assume that all the required Concepts are part of the SNOMED CT core concepts. The set of Concepts is shown in Table 6.6.

Table 6.6: Sample Concepts for Navigation Subset

Concept Identifier	Preferred Term
309902002	Clinical Oncology Department
55342001	Neoplastic Disease
362956003	Procedure/Intervention
108290001	Surgical Procedure
69960004	Radiation Oncology
309902002	Chemotherapy Regimen

In order to complete the navigation Subset we must simply add entries to the relational Subsets table, and to the Subset Members Table. For each new Subset an entry is required in the Subsets Table. The Subset Members Table uses the field LinkedID to specify the navigational children for each entry. The member status specifies the display order for the children.

Table 6.7: Subsets Table

Field Name	Value
SubsetID	309902022
SubsetOriginalID	
SubsetVersion	1.0
SubsetName	Clinical Oncology Department
SubsetType	Navigation
LanguageCode	
RealmID	
ContextID	

Table 6.8: Subset Members Table

SubsetID	MemberID	MemberStatus	LinkedID
309902022	309902002	0	55342001
309902022	309902002	1	362956003
309902022	362956003	0	48692003
309902022	362956003	1	108290001
30902022	362956003	2	6996004

Displaying a navigation subset

Displaying a navigation subset is very similar to displaying an “IS_A” hierarchy, as described in section [6.6.1]. Having mastered the techniques for using the “IS_A” relationship, the key to using navigational hierarchies lies in seeing which columns of the Subset Members table correspond to similar columns in the Relationship table.

For example, in the Relationship table we can say “Procedure by site” “IS_A” “Procedure,” this would be expressed in the Relationship table as follows:

ConceptID1	RelationshipType	ConceptID2
362958002	116680003	71388002

In the navigation Subset Members table the equivalent would be “Radiation Oncology” (is a navigational child in Subset) “Clinical Oncology Department (of) “Procedure/Intervention.”

LinkedID	SubsetID	MemberID
108290001	309902002	362956003

Clinical Oncology Department
- Neoplastic Disease
- Procedure/Intervention
  - Surgical Procedure
  - Radiation Oncology
  - Chemotherapy Regiment

Figure 6.6: Navigation hierarchy example: Clinical Oncology navigation Subset

Completeness of navigation hierarchies

There is no particular requirement for navigational Subsets to contain links that will allow access to all descendants of a Concept included in the navigational Subset. Implementers of SNOMED may choose to display only those Concepts they have included in the navigation Subset Members table. Alternatively they may wish to use the navigation Subset to organize the top-level hierarchy, but then switch to using the “IS_A” hierarchy from the Relationship table when navigation reaches a certain depth.

6.7 Supporting Post-Coordination

6.7.1 Access to qualifiers and refinable characteristics

A terminology server should enable an application to review the refinable defining characteristics and the specified set of qualifying characteristics for any selected Concept.

6.7.2 Deriving canonical forms

Introduction

An advanced terminology server should be able to report the short or long canonical form for any identified Concept or for any post-coordinated expression.

More details of the SNOMED Canonical Table are included in the documentation accompanying that table. Examples of use are given in various sections of this guide including SNOMED CT storage options for effective retrieval [7.3.3] and Appendix F. This section outlines the process of determining the canonical form of a Concept.

Note that the canonical form of a Concept may be refined in a new release and should always be computed from the most recent release tables.

Literal and extended canonical forms of primitive Concepts

Two views may be taken about the canonical form of a primitive Concept. The literal view says we do not know enough to fully define the Concept, therefore all we can say is that “it is what it is.” The alternative extended view accepts that it is not fully defined but expresses what is known about the Concept. These two views can both be derived directly from the distributed files.

The Canonical Table contains the proximal primitive supertype ancestors and any defining characteristics that distinguish the Concept from those ancestors for primitive Concepts as well as for fully defined Concepts. However, by recognizing the IsPrimitive field in the Concepts Table these incomplete definitions can be selectively ignored when retrieving canonical forms for a selected Concept.

A terminology service that returns the canonical form for a Concept should allow selection of either the “literal” or “extended” canonical data. If the “literal” form is requested, the canonical form returned should be an “IS_A” that references the selected concept.

Examples:

The primitive concept “Well person screening”

- “Literal” canonical form
- “IS_A” “Well person screening”
- “Extended” canonical form
- ““IS_A” “Procedure”
- “HAS INTENT” “SCREENING intent”

The “literal” canonical form is most useful when specifying a query.

Example:

A query for “*Well person screening*” should return all Concepts that are “IS_A” subtypes of “*Well person screening*.” This should include the Concept “*Well person screening*” itself.

The “extended” canonical form may be useful when testing a Concept in a record against some more general criteria. If we want all screening activities, then a form that makes this characteristic explicit is useful.

Short canonical form

The short canonical form consists of the primitive supertype ancestors of a Concept plus those defining characteristics of the Concept that differ from those of the primitive supertypes.

The short canonical form is computable from the Relationships Table and Concepts Table.

However, it is also more immediately available in the distributed Canonical Table using the following approach for the concept CPT_A:

If the “literal” version is required AND CPT_A.IsPrimitive=1

- The short canonical form is a Relationship that expresses following:
  - CPT_A ““IS_A” CPT_A

Otherwise

- The short canonical form is the collection of rows from the Canonical Table where
  - Canonical.ConceptID1=CPT_A.ConceptID

Long canonical form

The long canonical form is similar to the short canonical form. However, it includes all the defining characteristics rather than only those needed to distinguish the Concept from its proximal primitive supertypes.

Like the short canonical form, the long canonical form is computable from the Relationships Table and Concepts Table. However, it is much easier to derive it using the distributed Canonical Table and Relationships Table in the following way for the Concept CPT_A:

If the "literal" version is required AND CPT_A.IsPrimitive=1

- The long canonical form is a Relationship that expresses following:
  - CPT_A ""IS_A" CPT_A

Otherwise

- The long canonical form is the collection of rows from the Canonical Table where
  - Canonical.ConceptID1 = CPT_A.ConceptID
  - Canonical.RelationshipType = 116680003 ("IS_A")
- PLUS the rows from the Relationships Table where
  - Relationships.ConceptID1 = CPT_A.ConceptID
  - Relationships.RelationshipType **NOT** = 116680003 ("IS_A")
  - Relationships.CharacteristicType = 0 ("defining")

Canonical forms of post-coordinated expressions

The canonical form for a post-coordinated expression is the combination of the canonical form of the primary Concept(s) with the additional qualifiers and refined defining characteristics. If the post-coordinated expression refines a defining characteristic that is present in the canonical form of the primary Concept then the unrefined instance of that characteristic is removed from the canonical form.

6.8 Supporting Selective Data Retrieval

Introduction

This short section is concerned with specific terminology services that may be required to support selective retrieval. These services are additional to the foundation services related to testing of subtype subsumption and working with post-coordinated expressions and canonical forms.

A more detailed discussion of implementation of selective retrieval record services is provided in Section 7.4.

6.8.1 Creating queries

A terminology server should support the creation of queries that retrieve SNOMED CT encoded data by facilitating the generation of predicate statements.

For example, a terminology server may generate an SQL predicate list that includes the ConceptIDs of all unique subtype descendents of a specified Concept. Some constraints on this functionality may be necessary as top-level or other general Concepts may generate extremely long lists of descendent ConceptIDs.

6.8.2 Creating legacy queries

A terminology server should support the creation of queries that retrieve data encoded in SNOMED International (SNOMED version 3.x), Clinical Terms Version 3 or earlier versions of the Read Codes.

For example, a terminology server may generate an SQL predicate list that includes the SNOMEDIDs or CTV3IDs of all unique subtype descendents of a specified Concept. Some constraints on this functionality may be necessary as top-level or other general Concepts may generate extremely long lists of descendent identifiers.

6.9 Applying Subsets

6.9.1 Introduction

A Subset refers to a set of Concepts, Descriptions, or Relationships that are appropriate to a particular language, dialect, country, specialty, organization, user or context.

In its simplest form, a Subset is a list of SNOMED Clinical Terms Identifiers (SCTIDs). Each SCTID refers to one component of SNOMED CT, and makes it a member of the Subset (a Subset Member).

> **Example:**
>
> As an analogy, think of SNOMED CT as a book. A Subset is like an index entry pointing to a set of pages relevant to a particular topic.

Subsets may be used to derive tables that contain only part of SNOMED CT. In some cases, these cut-down "Standalone subset" tables may also be centrally distributed (e.g. a release table containing only Descriptions that are valid in a chosen language). The Subsets described in this document are additional tools for enhancing the usability of SNOMED CT.

Complete technical information on subsets, including the subset table design can be found in the Technical Reference Guide.

6.9.2 Subset Types

SNOMED CT supports the Subset Types listed in the following table.

Subset Type	Summary
Language Subset	The Descriptions applicable to a Language or Dialect. May indicate whether each Description contains the Preferred Term, Synonym, or Fully Specified Name for the associated Concept in that language or dialect.
Realm Concept Subset	The Concepts applicable for a particular Realm. May assign priorities to Concepts appropriate to the Realm.
Realm Description Subset	The Descriptions applicable for a particular Realm. May indicate whether each Description contains the Preferred Term or a Synonym used in that Realm.
Realm Relationship Subset (for future use)	The Relationships applicable to a particular Realm.
Context Concept Subset	The Concepts applicable to a particular Context Domain.
Context Description Subset	The Descriptions applicable to a particular Context Domain.
Navigation Subset	A set of Navigation Links representing an ordered hierarchy appropriate for display and user navigation.
Duplicate Terms Subset	Indicates the suggested priority among sets of Descriptions containing identical terms.

6.9.3 Subset Members

The format of the Subset Members Table appears quite simple, containing only 4 fields:

- SubsetID
- MemberID
- MemberStatus
- LinkedID

There are some underlying subtleties to this table however.

The LinkedID field is only valid for

- Navigation Subsets, where it specifies the navigational children of the subset member (see Section [6.6.4] of this document for more details).
- Duplicate Terms Subsets, where it refers to the Description which is most highly prioritized.

The field MemberStatus has a different significance depending on the type of subset specified in the Subset Table.

A brief description of the significance of the MemberStatus in each type of subset is given below.

Significance of MemberStatus in a Language or Realm Description Subset

The MemberStatus specifies the role of the reference Description in this Language or Realm. There are three roles.

1 =Preferred Term

2 = Synonym

3 = Fully Specified Name

Note that the following rules apply to each Subset:

- A Description cannot be assigned the role Fully Specified Name unless its DescriptionType is also "Fully Specified Name."
- Only one Description of each Concept may be assigned as the Fully Specified Name.
- If no Description of a Concept is assigned the role Fully Specified Name, the FullySpecifiedName in the Concepts Table assumes this role.
- A Description cannot be assigned the role Preferred Term or Synonym if its DescriptionType is "Fully Specified Name."
- One and only one Description of each Concept must be assigned the role "Preferred Term."
- Any number of Descriptions of a Concept may be assigned the role "Synonym."

Significance of MemberStatus in a Realm Concept or Context Concept Subset

The MemberStatus specifies the priority of the Concept within the Subset, with lower numbers denoting higher priority. The priority may be used to determine order of display. Concepts with priority status of "0" would appear at the top of the list.

Significance of a MemberStatus in a navigational Subset

The MemberStatus specifies the order of the child concepts within the navigational hierarchy. This may be used to determine the display order of children when displaying a navigational Subset.

Significance of MemberStatus in a Realm Relationship (for future use) or Context Description Subset

The MemberStatus has no particular significance in these types of subsets, and will be set to the value of "1," meaning include this table row in the subset.

6.9.4 Using subset definition files (future use)

Subset definition files are the preferred way of specifying the content of subsets. The subset definition file is a series of logical clauses, expressed as XML mark-up, which defines what should be included in the subset. While all subsets released by the IHTSDO are currently distributed in the relational file format, there are plans to provide these subsets as XML definition files. This method of specifying subsets has several advantages:

- It is less verbose as many members can be included by a single clause;
- It is likely to be resilient to the effects of applying updates and Extensions.
- One Subset can be used as the basis for another Subset. The defining clauses express differences between the base Subset and the defined Subset.

Example:

A UK English dialect Subset may be expressed as differences from an International English language Subset.

- Members can be added to a Subset according to hierarchical Relationships.

Example:

A single defining clause can add all "IS_A" descendents of the Concept "Musculoskeletal procedure" to an orthopedic specialist Subset.

Software developers who wish to use subset definition files will need to produce software that will generate relational subset tables from a subset definition. Please contact the IHTSDO for current information as this format evolves.

6.9.5 Using relational Subset tables

Any Subset Definition can be automatically converted into the relational form. However, the relational form is release version dependent whereas some Subset Definitions may be applicable to more than one release of SNOMED CT.

An application may use the relational form of a Subset in one of three ways:

- As part of a run-time query to retrieve, filter or sort members of a Subset.

Example:

An SQL query with a join between the Subsets Table and a core table could be used to:

- Provide a pick list of all members of a Subset.
- Filter a search or hierarchical list so that only members of the Subset are shown.
- Order a search or hierarchical list according to the sequence specified in the Subset.

- To pre-filter a SNOMED CT table when it is imported from a release file.

Example:

You might choose to exclude Descriptions in a language or dialect not required.

- Care should be taken when considering the use of pre-filtering with Concepts. If SNOMED CT data from other systems is received it will not be possible to interpret it unless the relevant Concepts (including Concepts used directly or indirectly in a definition) are present on the system.

- To modify the application's internal representation of Concepts or Descriptions to allow efficient application of the Subset:
  - Possible modifications include addition of flags or indexes or some type of pre-sorting of data.
  - Objectives of this include:
    - Optimizing frequently used Subsets
    - Facilitating rapid switching between different Subsets depending on context or user configuration.

Note the actual release file formats are tab-delimited files that allow the information to be imported to any relational database management system.

6.9.6 Practical uses of Subsets

Subsets can be used for many different purposes. This section outlines some of the ways in which the mechanism specified in this document can be used to meet different practical requirements. The uses outlined are illustrative examples and do not represent all possible applications of Subsets.

Controlling data entry

Many clinical applications include facilities for data entry to be controlled or assisted by protocols, templates or structured data entry forms. Different sets of candidate Terms or Concepts may be appropriate to each data entry field. The sets of candidate Terms or Concepts for a field may be very large (e.g. any operative procedure) or very small (e.g. the possible observations from a particular examination).

The Subset mechanism allows Concepts relevant to particular uses to be specified as follows:

- A Context Concept Subset or Context Description Subset can be specified for each data entry field.
- A Context Concept Subset refers to all the Concepts that are applicable to a particular category of use.

  - A Context Concept Subset allows any Description associated with the set of permitted Concepts to be used (i.e. a Preferred Term or a Synonym).
- Context Description Subset refers to all the Descriptions that are applicable to a particular field.
  - A Context Description Subset only allows a specified set of Descriptions to be used. The permitted set of Concepts is implicitly any Concept associated with permitted Description.

A SNOMED CT Enabled Application can make use of Context Subsets if it is able to associate data entry fields with an identifiable Context Domain. There are several alternative ways of using Context Subsets. These include:

- For fields associated with a Context Subset containing only a small number of permitted values, the members of the Subset may be used to populate a drop-down list or another similar option selection tool.
- For fields associated with a larger range of options, a Context Subset may be used as a filter for searches.
- When using automated text encoding tools a Context Subset may limit the scope of possible encoding and thus should accelerate and improve the accuracy of parsing.

Using terminology in electronic health records

When a user searches SNOMED CT, they should only see Descriptions associated with Concepts that are usable in the context in which they are working. The Concepts that form part of SNOMED CT have various different uses. There are many examples of Concepts that are meaningless if used in an inappropriate context. These include:

- Concepts that are used purely to structure and organize the terminology but which would be meaningless in a patient record (e.g. the root Concept, top-level Concepts).
- Concepts that may be useful as headings for sections in a record but which convey no clinical meaning (e.g. general Concepts like “Laboratory results,” “Medication”).
- Concepts that refer to an organism or material which are only meaningful in the context of other information (e.g. “Aspirin” is meaningful in a patient record only as part of, or in relation to, another Concept such as “Overdose,” “Allergic reaction” or “Prescribed”).
- Concepts that refer to disorders are meaningless or misleading if used to record a procedure.

The Subset mechanism allows Concepts relevant to particular uses to be specified as follows:

- A Context Concept Subset can be specified for any distinct category of use associated with a set of Concepts.
- A Context Concept Subset contains a RealmID and ContextID which identify the Context Domain to which it applies.
- A Context Concept Subset refers to all the Concepts that are applicable to a particular category of use.
- A Context Concept Subset only refers to Concepts. However, it also affects the availability of Descriptions. A Description associated with a Concept that is not used in a particular context is not accessible.

A SNOMED CT Enabled Application should be able to recognize usage categories that are relevant to it and should apply the appropriate Context Concept Subsets to:

- Restrict access to Concepts (and associated Descriptions) so that only those referred to by the Context Concept Subsets are accessible.

Languages and Dialects

SNOMED CT is designed as a multilingual terminology.

The Subset mechanism enables SNOMED CT to support different Languages and Dialects in the following way:

- There is a Language Subset for each supported Language and Dialect.
- The LanguageCode identifies the Language or Dialect.
- A Language Subset refers to all the Descriptions that contain Terms expressed in that Language or Dialect.
- Each member of a Language Subset has a MemberStatus. This indicates the use of the Description in relation to its associated Concept. In each Language or Dialect, a Description may represent a:
  - Preferred Term
  - Synonym
  - Language-specific Fully Specified Name

A SNOMED CT Enabled Application should be able to:

- Allow a selection of a particular Language Subset as a configuration option.
- Restrict access to Descriptions so that only those referred to by a selected Language Subset are accessible.
- Treat the Preferred Terms and language-specific Fully Specified Name referred to by a selected Language Subset in a manner appropriate to their specified usage.

Some applications may also support multiple Languages or Dialects allowing selection of combinations of Language Subset.

Managing duplicate terms

A Fully Specified Name in a given Language or Dialect is uniquely associated with one and only one Active Concept. However, this requirement does not apply to other Terms (i.e. Preferred Terms and Synonyms).

SNOMED CT permits the same Term to occur in more than one Active Descriptions each associated a different Active Concept. This enables Preferred Terms and Synonyms to be phrased in the natural languages of clinicians without discriminating between those in different disciplines or specialties who use the same words or phrases to mean different things.

Example:

- "Liver" – may refer to a foodstuff (dietician) or to a body organ (surgeon)
- "Fundus" – may refer to part of the eye, stomach or uterus depending on specialty and context.
- "Aspiration of stomach content" – may refer to a procedure (following poisoning) or a complication of anesthesia or childbirth.
- "Hypertension" – may refer to a finding (raised blood pressure) or to a recognized diagnosis with particular diagnostic criteria.

Similarly the same words or phrases may be used to refer to the Concept with different levels of precision.

Example:

"Arm" – may refer to the entire arm (as in "amputation of arm") or to part of the arm (as in "injury to the arm").

The existence of multiple Descriptions containing exactly the same Term but linked to different Concepts does not cause any technical problems but may cause confusion at the user interface. The appearance of several identical phrases in a search list presents a dilemma and applications need to present sufficient information to allow the user to choose between them. In most cases this requires the application to provide a view of the Fully Specified Names associated with any Duplicate Terms.

The Duplicate Terms Subset, part of the SNOMED CT Developer Toolkit, is intended to assist applications to identify and manage the presentation of Duplicate Terms by:

- Identifying the Duplicate Terms in a particular release of SNOMED CT.

  This simplifies the processing required to determine if a duplicate exists and will require disambiguation.

- Indicating a priority between Duplicate Terms.

  Applications may use these priorities to support sensible display ordering or default selection.

The Duplicate Terms Subset is Language and Dialect dependent since the occurrence of and priorities between duplicates will vary according to the applicable set of Terms specified in a Language Subset. In addition, specialty or user specific priorities may be applied by alternative Duplicate Terms Subsets derived from the distributed files.

- Each current Description containing a Term that is also present in another current Description is a member of the Duplicate Terms Subset.
- The Subset Member has a status that indicates its priority.
- The LinkedID refers to the Description which contains the same Term and has the highest priority.

  This reference links together a set of Descriptions which contain the same Term.

Specifying "natural" display orders

The formal subtype Relationships of SNOMED CT are specified as logical definitions. These provide a powerful tool for organizing, retrieving and aggregating data. These subtype Relationships can also be used for Navigation to locate a more or less specific Concept. However, the requirements for Navigation differ from those for logical definition. The following factors affect the efficiency of Navigation but have no relevance to logical definitions of Concepts:

- Frequency of use of Concepts and their relative accessibility
- Depth of nesting of a hierarchy
- Length of a displayed hierarchy level
- Order of listed Concepts
  - Matching a natural order (e.g. cranial nerves, cervical vertebrae, etc.)
  - Reflecting frequency of use

- User association of ideas that are not linked as subtypes of one another.

The Relationships Table does not specify the order in which Relationships should be displayed and places no limits on the size or depth of any hierarchy. There are several reasons for this:

- The display order of many Relationships is arbitrary.
- Different display orders may be appropriate to the same Relationships depending on user preferences and the context.
- The display order of Relationships is not fundamental to an understanding of the related concepts. A change in display order does not alter the meaning of a Concept. Therefore, it need not be subject to the rigorous change control applied to the core tables.

There is a requirement for one or more Navigation hierarchies, which are configurable without undermining the logical definitions of SNOMED CT.

The Subset mechanism includes Navigation hierarchies, which allow a natural or preferred display order to be defined for a set of Navigation Links. They are represented as follows:

- A Navigation Subset specifies an ordered list of Navigation Links for Concepts that act as nodes in the hierarchy.
- Each Navigation Link is represented as a member of the Navigation Subset
- Each Navigation Link specifies the ConceptIDs of a parent Concept and a Navigation child Concept and an integer used to order the list of children.
- A Navigation hierarchy may include both ordinary SNOMED CT Concepts and special Navigation Concepts:
- Navigation Concepts exist only for the purpose of Navigation and are not suitable for recording or aggregating information.
- This allows a Navigation hierarchy to contain nodes which represent user recognizable composite Concepts that are not part of the formally defined content of SNOMED CT.
- A Navigation hierarchy may be similar to the subtype hierarchy specified by "IS_A" Relationships but it differs in the following ways:
    - Navigation Links need not represent logical semantic relationships between Concepts. They can be used to link to different Concepts that a user may seek in an indirectly associated part of the SNOMED CT hierarchy.
    - Navigation Links need not include a complete set of subtypes for a Concept. Instead they may be limited to those most commonly used.
    - Navigation Links can skip levels in the subtype hierarchy to show commonly used Concepts to a higher more readily accessible level.
    - Navigation Links can use Navigation Concepts to add intermediate levels into the formal subtype hierarchy to rationalize screen displays.
    - Navigation Links are ordered and thus provide added value even when they have a one-to-one relationships with the subtype hierarchy (for example, while the "IS_A" Relationships identify all the cranial nerves a set of Navigation Links can order them in the expected order 1 to 12).

A SNOMED CT Enabled Application should support ordered hierarchical displays based on the Navigation Links in a selected Navigation Subset. SCT Enabled Applications should also provide access to the subtype hierarchy so that a complete set of subtypes can be accessed where required. A Navigation display should therefore always include an option to switch to a subtype view of the same focus Concept.

National requirements for specific Concepts

SNOMED CT is designed for use in many different countries and consequently includes a certain number of country-specific Concepts. A country, such as the United Kingdom for example, may have specific requirements for the representation of Concepts that are not meaningful in the United States or other countries. These variations are particularly significant for the interfaces between clinical care, service administration and reimbursement. National laws and conventions may also create additional refinements of more general Concepts.

A user or group of users usually only requires access to Concepts deemed to be relevant in the country in which they are working.

The Subset mechanism meets the requirement for national and region variations in the use of Concepts in the following ways:

- Realm Concept Subset can be specified for each country in which SNOMED CT is used.
- If a country has administrative regions with different requirements, these can also be specified with a Realm Concept Subset.
- A RealmID identifies the country or region.
- A Realm Concept Subset refers to all the Concepts used in that country or administrative region.
- A Realm Concept Subset only refers to Concepts. However, it also affects the availability of Descriptions and Relationships. A component that refers to a Concept that is not used in a Realm is implicitly also excluded from use in that Realm. See also National, Organizational and Specialty differences in Terminology [6.9.6].

A SNOMED CT Enabled Application should be able to:

- Restrict access to Concepts so that only those referred to by a selected Realm Concept Subset are accessible.
- Restrict access to Descriptions so that only those associated with Concepts referred to by a selected Realm Concept Subset are accessible.
- Restrict use of Relationships so that any Relationship that refers to a Concept that is not referred to by the Realm Concept Subset is rendered unusable.

Example:

- An oncology subspecialty organization Realm Concept Subset may refer to the hierarchical descendants of a general Concept representing "Grant funding for cancer research"
- A UK NHS Realm Concept Subset may refer to the hierarchical descendants of a general Concept representing "NHS National Service Frameworks."

Regional variations in disease prevalence

There are substantial differences in the prevalence of diseases in different regions in which SNOMED CT may be used. Users will expect to find the conditions they commonly deal with, without being distracted by long lists of conditions they rarely see.

In the main body of SNOMED CT, all Concepts are equally visible irrespective of their relative prevalence. The Subset mechanism can be used for prioritizing the Concepts in the following way:

- Realm Concept Subset can be specified for a region with a particular pattern of disease prevalence.
- A RealmID identifies the region.
- A Realm Concept Subset refers to all the Concepts used in that region. For each of these Concepts, the Subset may specify a MemberStatus that assigns a priority to that Concept. Concepts that are used most frequently in a region are assigned first priority (MemberStatus=1). There are nine priority levels (MemberStatus=1 to 9). The least frequently used Concepts are assigned the lowest priority (MemberStatus=9).
- A Realm Concept Subset only refers to Concepts. However, the assigned priority also affects the Descriptions associated with those Concepts.

A SNOMED CT Enabled Application should be able to:

- Prioritize access to Concepts relative to one another according to the MemberStatus specified by a selected Realm Concept Subset.
- Prioritize access to Descriptions relative to one another according to the MemberStatus assigned to the associated Concepts by a selected Realm Concept Subset.

The way in which access is prioritized depends on the nature of the application and its operating environment. However, examples of prioritization include:

- Showing Descriptions associated with high priority Concepts before those with lower priority when searching for word or phrases.
- Showing Concepts with high priority before their less highly prioritized siblings in hierarchical displays.
- Initially listing Concepts and associated Descriptions with priority above a specified threshold and requiring an additional step to access those assigned lower priority.

Specialty and discipline-dependent variations in use of Concepts

SNOMED CT contains Concepts used by many different groups of health professionals. The frequency of use of these Concepts depends on the professional discipline and/or clinical specialty of the user. It is important to ensure that the user is able to access the Concepts that they use frequently, without being distracted by thousands of textually similar Concepts they rarely require.

The Subset mechanism can be used for prioritizing the Concepts in the following way:

- A Realm Concept Subset can be specified for each clinical specialty and/or discipline.
- A RealmID identifies the specialty.
- A Realm Concept Subset refers to all the Concepts used in that specialty. For each of these Concepts, the Subset may specify a MemberStatus that assigns a priority to that Concept. There are nine priority levels from 1 (highest priority) to 9 (lowest priority).
- A Realm Concept Subset only refers to Concepts. However, the assigned priority also affects the Descriptions associated with those Concepts.

Local needs of organizations or individual users

The previous sections have dealt with requirements of countries, regions and specialties. Organizations and individual users may also have similar requirements for restricting or prioritizing access to particular Concepts.

The Realm Concept Subsets described in the previous sections may also be applied at a more local level to meet the needs of an organization, user or group of users.

In practice, a local organization or individual user will often need a Realm Concept Subset derived by merging or refining other Realm Concept Subsets.

Example:

A cardiologist, working in "Any Town General Hospital," an NHS hospital in the UK, may be best served by a Subset derived from:

- A combination of the existing Realm Concept Subsets specified for:
- UK NHS
- Cardiology specialty
- Any Town General Hospital
- Modifications to take into account any local factors (e.g. special research interests of the local cardiology department or clinician).

National, organizational and specialty variations in term preferences

In addition to requirements for different Concepts, a country, organization or specialty may have different preferences for the Terms to be used to describe identical Concepts.

The Subset mechanism meets the requirement for national and region variations in the use of Terms in the following way:

- Any Realm that requires a specialized refinement of a Language or Dialect may specify a Realm Description Subset.
- A Realm Description Subset is similar to a Language Subset but differs in two ways:
  - It contains a RealmID that identifies a Realm within which it applies.
  - It must not be used to override the language-specific Fully Specified Name role assignments made by a Language Subset.
- Support for Term variations is Language or Dialect dependent. Therefore, in a multilingual country or organization, there may be a single Realm Concept Subset but several separate Realm Description Subsets.

A SNOMED CT Enabled Application should be able to:

- Use a selected Realm Description Subset, in place of a Language Subset, to restrict access to terms and to assign the Preferred Term and Synonym roles to different Descriptions.

Managing the coded content of messages

Healthcare messages include fields that can be populated with codes from clinical coding schemes. SNOMED CT provides Concept identifiers as a means of encoding Concepts. These concept identifiers are suitable for use in appropriate fields of many clinical messages.

Implementations of clinical messaging typically constrain the range of values that can be applied to particular fields. There are several reasons for this:

- To ensure that the information encoded is meaningful as a value for the specified field.

Example:

A field that is intended to describe the nature of investigation may contain a code that means “Serum glucose measurement” but should not contain a code that means “Hypoglycemia.”

- To ensure that receiving application is able to process the message.

Example:

A locally added code value may be valid in a particular application but should not be used if the receiving application needs to retrieve, process or analyze the coded part of the message.

- To ensure adequate detail and specificity.

Example:

A field used to report an operative procedure could contain a code for “Abdominal procedure.” However, this would not be adequate to meet the business purpose served by a message.

- To avoid unnecessary detail or diversity.

Example:

A biochemical investigation could be reported using a code that represents various detailed aspects of the method used to perform the investigation. Such details may be unnecessary to a clinician and may complicate the analysis, charting and graphing of a series of results reported at different levels of detail.

Practical examples of these constraints are:

- The UK NHS “bounded-lists” of codes that are permitted for use in specified fields of the pathology report message.
- The HL7 specification of “Vocabulary Domains” associated with every coded field in HL7 Version 3 messages.
- The members of the SNOMED CT US Drug Extension that contains pharmaceuticals approved for distribution in the United States.

The Subset mechanism allows Vocabulary Domains (or bounded-lists) of SNOMED CT Concepts to be specified in the following way:

- A Context Concept Subset is specified for each Vocabulary Domain.
- A Context Concept Subset lists the SCTIDs of all Concepts applicable to the Vocabulary Domain associated with a particular message field.

Where appropriate, a SNOMED CT Enabled Application should be able to use Context Concept Subsets:

- To generate messages that contain SCTIDs representing SNOMED CT Concepts appropriate to particular message fields.

- To validate messages to ensure that no fields contain SCTIDs that are inappropriate to the relevant Vocabulary Domain.

6.10 Terminology Server Software

6.10.1 Introduction

This section outlines the possible characteristics of software that provides terminology services through a programmable interface. Such software represents an approach to development that may enable more rapid implementation of SNOMED CT.

This document does not specify a particular Application Programming Interface (API) for accessing SNOMED CT services. Instead it sets out the general principles and options for delivery and use of a terminology server[9].

6.10.2 Terminology server functionality

A terminology server should be able to deliver all the essential terminology services identified in this section. It should also provide the recommended terminology services and should achieve a performance that meets the more general requirements for the functionality of SNOMED CT Enabled Applications.

Terminology server may provide two types of service:

- Reference Services (see Figure 6.7)
  - Services that do not include a user interface.
  - The client application may use reference services to undertake many different functions.
  - For some of these functions the client application will populate an appropriate user interface component.

Example:

A reference server may return a list of Descriptions matching a particular search string. The client application may use this data to populate a list from which a user makes a selection.

- User Interface (UI) Services (see Figure 6.8)
  - Services that include the one or more user interface components that can be used in and programmatically accessed by the client application.

Example:

A UI server may provide a control that includes a text box and a list. When the user types in the text box, the server populates the list and allows the user to select an item. The selected item is accessible from the client program.

  - One possible type of UI service is a SNOMED CT browser with an API for returning selected data to a client application.
    - This may be useful as mechanism for providing some SNOMED CT capabilities to an application. However, it is less suitable for frequent entry of SNOMED CT encoded information.

---

[9] A discussion draft of a SNOMED CT API is available for review and pilot implementation.

Figure 6.7: Client application with a Terminology Reference server

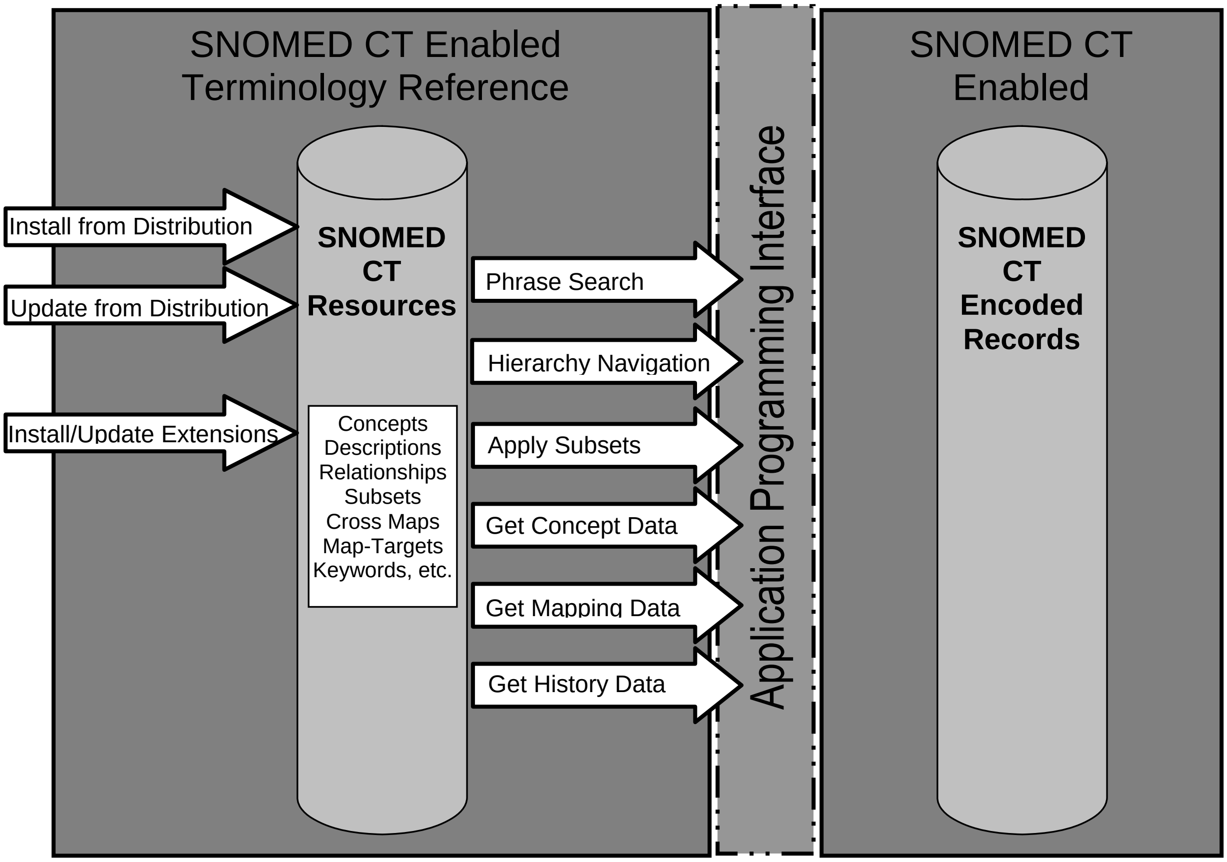
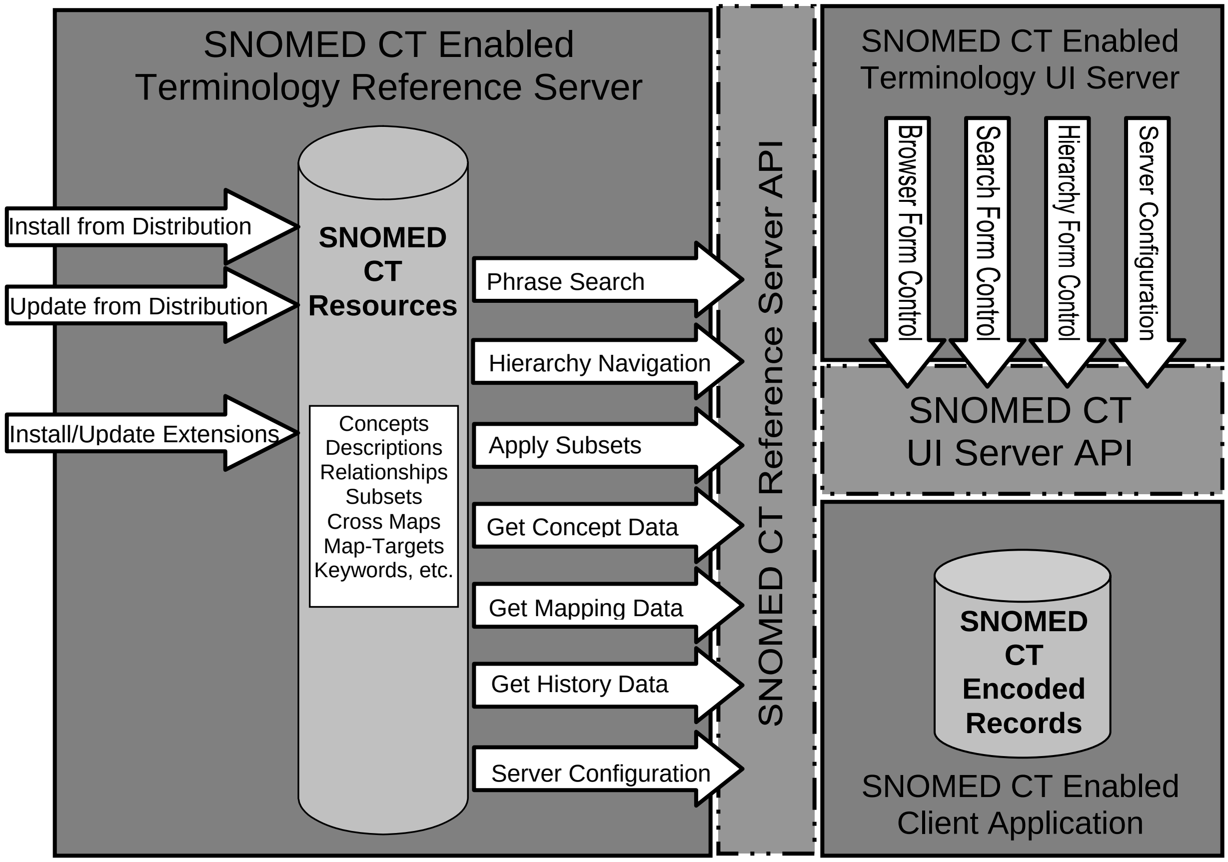


Figure 6.8: Client application with Terminology Reference and UI servers.



6.10.3 Programming interface technology

As noted earlier this document does not specify a particular API. The general principles outlined in this guide may be delivered using various types of interfaces as prescribed by various different technologies. Potential technologies include:

- Active-X / COM / DCOM in Microsoft Windows environments
- CORBA
- Various web-based technologies

Decisions on which technologies to support depend on the intended functionality, performance, accessibility, ease of use and support requirements for maintenance or updates.

7 Record Services Guide

7.1 Introduction

The primary use of SNOMED CT is to enable information to be entered in a health record and stored in a manner that enables selective retrieval. Effective mechanisms of selective retrieval are required to support aggregation, analysis and decision support. Information in a health record may need to be communicated in the interests of the patient or to enable larger scale aggregation and analysis. Communication needs preserve the semantics of the information expressed using SNOMED CT concepts. The following sections discuss requirements for record services that support entry, storage, retrieval and communication of SNOMED CT encoded information.

7.2 Entering SNOMED CT Information

7.2.1 Introduction

Various methods may be used to enter SNOMED CT information into a record. These employ one or more of the terminology services specified in the previous section of this guide.

7.2.2 Using text searches and subtype hierarchy navigation

Selection in a browser

The starting point for a consideration of data entry is an efficient method for performing text searches and subtype hierarchy navigation. When these functions are integrated in a terminology browser it is possible to select a Concept by text search and then to refine or generalize the selection to identify a more appropriate Concept for recording.

A terminology browser built into an application or offering a programmable interface can be used to allow a user to select a Concept (and/or Description) and enter this into a record.

This method of data entry allows unconstrained selection of any Concept from SNOMED CT. This can sometimes be useful but such an unrestricted method should only be used as a fallback, when more selective approaches cannot be used.

Limitations of simple browsers

A general-purpose browser capable of searching and navigating through the SNOMED CT hierarchy is a simple starting point. However, this approach is unlikely to meet the requirements for anything other than occasional entry of SNOMED CT encoded information. More selective mechanisms tailored to particular data entry contexts are likely to be more usable and may promote more consistent data recording.

In most situations in which clinical data is entered, access to the full content of SNOMED CT through a simple search and hierarchy browser is unlikely to be necessary and may be cumbersome and unhelpful. The main reason for this relates to the size and structure of the terminology. As a result:

- Many terms may match a single word or short phrase resulting in a long list of options.
- The depth and breadth of the subtype hierarchy and navigation may require selection of choices from several screens to locate the required Concept.

There are many ways to improve and simplify SNOMED CT data entry. Some of these can be used in a wide range of situations. Others are specific to constrained contexts that occur in structured data entry driven by a template or protocol.

7.2.3 Optimizing searches for data entry

Extending searches and limiting duplication

The terminology services guide addresses ways of:

- Extending text searches to include similar words and phrases by making use of the Word Equivalents Table.
- Rationalizing text searches, which, in a simple search, return the same Concept more than once due to multiple matching Terms.

These techniques may be used to improve access to Concepts during data entry.

Searches with qualifier resolution

When typing text for a search, the user is unlikely to know if their intended entry can be represented by a single Concept or requires a post-coordinated expression involving additional Concepts or qualifiers. Where searches fail to find a pre-coordinated match, expansion of the search to support appropriate or commonly used qualifiers is likely to enhance usability.

Some terminology servers may provide a general facility of this type. Alternatively, a limited facility for recognizing commonly qualifying words may be used. For example, words such as “left,” “right,” “routine,” “emergency” and “severe” are applicable as qualifiers when not included in a pre-coordinated Concept.

Real time searching

Conventional text searches require the user to decide how many words to enter and then explicitly request a search. When a search fails to find any matches or returns a very long list of matches, the user is obliged to repeat the process. The need to undertake this type of user interaction for every coded entry is likely to create a significant disincentive to effective data entry.

One possible solution to this is an interface that performs real-time checking of the number of matches as the user types. The interface may indicate this to the user, allowing them to decide when to stop typing and commence the search. A further enhancement is to automatically return the list of matches whenever the user stops typing, or when the number of matches reduces to an acceptable level.

Background encoding

Techniques that support real-time searches and qualifier resolution may also be extended to enable background encoding of complete sentences as they are entered. This method can be applied to text entered by typing or by voice recognition.

As text is entered, the search mechanism attempts to narrow the selection. If this process eventually finds a single good match, this is used to encode the text. The match should be displayed allowing the user to override it, but the default action is to accept the encoding. If at the end of a sentence there are multiple possible matches, then these are presented for user selection.

There are many possible variants on this technique. For example, as the possible matches are narrowed down, the system could offer an auto-completion option similar to that used in web browsers and word-processors.

Automatic and semi-automatic encoding

Techniques similar to those used for background encoding can be applied to previously entered text or to text entered by voice recognition or optical character recognition. Where such methods are used there is likely to be a need for manual intervention to resolve uncertain encoding. The requirement for manual intervention will depend on the sophistication of the matching techniques and the extent to which accuracy is safety-critical. If encoded data is to be used by clinical decision support protocols, which may influence the treatment of a patient, extreme care is needed when using automatic encoding and tools that allow manual review are essential. A less rigorous approach may be acceptable where the purpose of encoding is for aggregation and analysis of large volumes of population data.

Mnemonics and personal favorites

Groups of people, such as practitioners of a discipline or specialty, frequently use similar sets of Descriptions and Concepts. Lists of widely understood (or easily learned) abbreviations or mnemonics that allow rapid entry of these commonly used concepts are recommended as a way of accelerating repetitive recording.

A similar facility may also be useful for individual users or organizations that have sets of Descriptions and Concepts that they use frequently. An easy way to use options to store and recall personal favorites with user-defined abbreviated access terms will enhance usability and significantly increase the speed of data entry.

User guidance may be necessary to minimize the risk of shortcuts such as these being overused. Unless the general search facilities are also easy to use, it is likely that users will favor the shortcuts even when it would be more appropriate to use a more accurate but less accessible Concept. An unchecked bias towards easy to record Concepts may lead to deterioration in data quality, statistical anomalies, and in the worst case, inappropriate treatment.

7.2.4 Constraining searches for data entry

Constraining searches by status

Searches should usually be filtered according to the status of the Concept and/or Description. An application should only allow Active Concepts and Descriptions to be entered in a patient record. Active Concepts and Descriptions include those with the status “current,” “limited” or “pending move.” Concepts and Descriptions with the status “limited” should also be excluded unless there is a specific requirement for access to classification of Concepts.

There are a few cases where a user may legitimately wish to search Inactive Concepts and Descriptions. Possible reasons for this include creating or editing queries that locate previously entered data recorded using Concepts and Descriptions that are no longer recommended for active use. Therefore, it must be possible to disable or vary the status filters applied to searches.

Constraining searches by subtype ancestors

Searches may usefully be limited to Concepts that have a specified supertype ancestor, which is appropriate for the context of a particular field, template or protocol.

Example:

When attempting to record the diagnosis "renal calculus," it is not helpful for a search to include the procedures that may be carried out to treat a renal calculus.

Constraining searches by subsets

Searches for Descriptions or Concepts may need to be constrained by Subsets. Applications should allow searches to be filtered, ordered or otherwise prioritized in accord with one or more active Subsets. Specifically, the search mechanism should support the following functions with respect to the following types of Subsets:

- Filtering of search and navigation results to include only those Descriptions that are referenced by the Language Subset may be applied to limit a search to those Descriptions applicable in a particular language or dialect.
- A Realm Concept Subset or Realm Description Subset may be used to filter, sort or highlight the results of text search or hierarchical navigation. This may simplify or encourage selection of Concepts or Descriptions used in a particular country, organization or specialty.
- A Context Concept Subset or Context Description Subset may be used to specify or order the valid Concepts for entry in a particular field.
- A Duplicate Terms Subset for the appropriate Language or Dialect may be used to identify descriptions containing the same Term and to enable appropriate selection by:
  - Using the prioritization and defaults specified by the MemberStatus and LinkedID.
  - Enabling the user to view and choose between the alternative Concepts by displaying the Fully Specified Name and/or defining characteristics.

7.2.5 Constraining and extending hierarchical navigation for data entry

Using the subtype hierarchy for data entry

The most visible hierarchical construct in SNOMED CT is the subtype hierarchy. This is constructed using a set of logical rules. The purpose of this hierarchy is to support data retrieval and aggregation by addressing the question "is concept-A a subtype of concept-B."

The same hierarchy can be used for data entry navigation but it is not designed for this purpose. Its depth and breadth are determined by logical rules of subsumption rather than by usability. As a result:

- There is no upper limit on the number of subtypes a Concept may have. This is true because there is no rule that determines the number of subtypes that a real world concept may have. However, long lists of options are not conducive to effective data entry.
- There is no fixed limit to the number of hierarchical steps between a generalized Concept and its most refined subtype. This is true since there is no preordained limit on the extent of possible refinement of a real world concept. However, data entry procedures that involve stepping through several levels of choices before reaching the required selection impair usability.

- The subtypes of a Concept do not have any particular order. The "is a" Relationship is primarily a property of the subtype Concept and does not express an ordinal position. This is true because logical subtypes are inherently an unordered set. However, a user is likely to find it easier to locate their required selection if members of hierarchical lists are displayed in some recognizable order.
- The issues of depth, length and order noted above are also subject to change between releases. The addition of an intermediate Concept or reclassification after the addition of new defining characteristics will introduce new layers in the hierarchy. Some Concepts will then move from the list of immediate subtypes of a Concept to become subtypes of a more refined Concept. Hierarchical changes may sometimes simplify navigation by reducing the number of choices at a given hierarchical level. However, the general effect of improvements in the subtype hierarchy will be to increase its depth and thus to increase the number of steps from a particular general Concept to its most refined subtypes.
- The nature of a subtype hierarchy means that there may be many routes from a given Concept to its more general descendants. This means that some of the choices presented for user selection are redundant since they simply offer alternative routes to the same Concept.

Routine use of subtype hierarchy navigation is not recommended for data entry. However, despite the drawbacks listed above, the subtype hierarchy may be useful for undertaking an exhaustive search for a particular refined Concept.

Example:

The Concept "Laparoscopic emergency appendectomy" can be reliably located by subtype navigation from any of its supertypes: "appendectomy," "laparoscopic appendectomy" or "emergency appendectomy."

Using Navigation Subsets to support data entry

Navigation Subsets provide alternative hierarchical representations of SNOMED CT. They are intended to support data entry by addressing the limitations of the subtype hierarchy discussed in the previous section.

- Usability constraints can be placed on the number of levels in the hierarchy and the number of options displayed at each level in the navigational hierarchy.
  - If there are relatively few options and many layers, the most common options can be brought to a higher level.
  - If there are long lists of options, these may be subdivided with less frequent options moved to lower levels.
  - Options that are rarely or never used by a particular user community can be excluded from a navigational hierarchy to limit the range of choices. According to requirements, these options may remain accessible by switching to a subtype view.
- Options at each hierarchical level can be ordered to meet the expectations of users and/or to facilitate rapid access to commonly used options.
- The available options at a particular level can be kept stable across releases without affecting the accuracy of the subtype hierarchy.

A Navigation Subset may be based on the foundation of the SNOMED CT subtype hierarchy. This can then be modified to add ordering and other features discussed above. An alternative starting point is a hierarchy of classification derived from another coding scheme or classification.

Note:

A Navigation Subset derived from the Clinical Terms Version 3 hierarchy is provided with the SNOMED CT Developer Toolkit as an example.

Alternative Navigation Subsets can be created from scratch (or as variants of a common source hierarchy) to provide views to support users with different requirements. Since navigational hierarchies do not affect interpretation, retrieval and aggregation, data entered in using different views can be analyzed consistently.

7.2.6 Constraining data entry

Introduction

Some Concepts or Descriptions displayed by searches, hierarchical navigation or other methods of data entry may not be suitable for recording in a patient record. Various reasons for this are discussed in the following sections. They include:

- ConceptStatus
- DescriptionStatus
- Special Concept
- Subtype relevance
- Subset inclusion
- Cross Mappability
- Context

Constraining data entry by Concept Status

Inactive Concepts should not be added to a record. Inactive Concepts include those with the following ConceptStatus values:

- **Retired:** The Concept should not be used. No further information is available.
- **Duplicate:** The Concept was found to be the same as another Concept. The duplicated Concept can be identified by following "SAME AS" Relationship and this may be used.
- **Outdated:** The Concept that is no longer meaningful due to changes in accepted understanding of biology, disease, health or related subject areas.
- **Ambiguous:** The meaning of the Concept is ambiguous because its associated Descriptions have different meanings. The Concepts representing possible alternative meanings can be identified by following the "MAY BE A" Relationships and one of these may be selected for use.
- **Erroneous:** There is an error in the representation of the Concept. The corrected Concept can be identified by following the "REPLACE BY" Relationship and this may then be used.
- **Moved Elsewhere:** The Concept is now maintained in a different namespace. The target namespace can be identified by following the "MOVED TO" Relationship. If release data for this namespace is available the relevant Concept can be identified by locating a "MOVED FROM" Relationship in the target namespace that refers to

the original Concept. The Concept from the target namespace may then be used. If the release data for the target namespace is not available, the Concept should not be used.

Active Concepts include those with different ConceptStatus values and depending on circumstances, some of these may not be appropriate to add to a record:

- **Current:** Suitable for general use.
- **Limited:** Only suitable for use if there is a requirement to record classification Concepts with limited semantic stability (e.g. terms that contain "not otherwise specified," "NOS," "not elsewhere classified," "NEC").
- **Pending move:** Suitable for use unless and until the Concept is available in another namespace. Concepts with this status will generally become inactive with the status "Moved elsewhere" (see above) in a subsequent release.

Constraining data entry by Description Status

It is not appropriate to allow Inactive Descriptions to be added to a record. Inactive Descriptions include those with the following DescriptionStatus values:

- **Retired:** These should not be entered in a record. No further information is available.
- **Duplicate:** The Description contains the same Term and the same ConceptID as another Description[10]. The duplicated Description should be identified either
    - By following the appropriate Reference.
    - By string matching among the other Descriptions of the associated Concept.
- **Outdated:** A Description that contains an obsolete Term which refers to a valid Concept. Another Description associated with the Concept should be used.
- **Erroneous** – There is an error in the representation of this Description. The corrected Description can be identified by following the appropriate Reference and this may then be recorded.
- **Inappropriate:** The Term in this Description should not be associated with this Concept. The same Term associated with its appropriated Concept can be identified either:
    - By following the appropriate Reference
    - By searching for the same Terms associated with as Active Description. More than one Active Description may be associated with a Term if that Term is used to describe more than one Concept.
- **Concept retired:** The Description is associated with an Inactive Concept. The Description was active until the Concept was inactivated. The Description should not be used and is retained only to provide valid Descriptions associated with the Concept and any legacy data previously recorded using this Concept.
- **Moved elsewhere:** The Description is now maintained in a different namespace. This applies where a Concept has also been moved to the new namespace (see above).

[10] Note that Duplicate Terms associated with different Concepts may be Active Descriptions and these are not marked with the DescriptionStatus "duplicate." See also Duplicate Terms Subsets.

Constraining data entry to exclude Special Concepts

Special Concepts are Concepts that are subtypes of the top-level Concept "Special Concept."

- **Inactive Concepts:** are subtypes of the Special Concept "Non-current Concept."
  - As noted in [7.2.6] Inactive Concepts should not be added to records.
- **Navigation Concepts:** are subtypes of the Special Concept "Navigation Concept."
  - Navigation Concepts should not be added to a patient record. These Concepts are only intended to rationalize and assist hierarchical navigation.
- **Namespace Concepts:** are subtypes of the Special Concept "Namespace Concept."
  - Namespace Concepts should not be added to a patient record. These Concepts are only intended to provide a reference point for "MOVED TO" Relationships when a Concept is moved to another namespace.

Constraining data entry according to subtype relevance

It may be necessary to prevent entry of a Concept in a subtype hierarchy that is inappropriate to a particular data entry field or to a particular part of a patient record. For example:

- An application should not allow a disorder Concept to be recorded in a field intended for recording a procedure (or vice-versa).
- An application should not allow a Concept that is a subtype descendant of the top-level Concept "attribute" to be recorded, except to associate another Concept with an appropriate qualifying value.
- An application should not allow a Concept that is a subtype descendant of the top-level Concept "qualifier value" to be recorded, except where it qualifies an appropriate attribute Concept.

Constraining data entry using Subsets

In some cases, identifying selected portions of the SNOMED CT hierarchy may be a sufficient constraint for entering data into a record. However, that is not always sufficient if Concepts from multiple hierarchies are required, or if there is a need to hone down the entry options from the full hierarchy. To meet these requirements applications should allow data entry to be constrained by Subsets.

Applications should be able to:

- Permit or prevent the entry of Concepts or Descriptions that are members of a specified Subset.

  Example:

  A UK GP system might:

  - Prevent the entry of Concepts in a subset that contains all Concepts that are non-human.
  - Enable the entry of Concepts in the "UK Administrative Subset" only when entering information in an administrative context.
- Encourage or inhibit the entry of Concepts or Descriptions according to their MemberStatus in a specified Subset.

Example:

A specialty system might prompt for confirmation when the user records a procedure not in a specified specialty Subset.

Constraining data entry based on cross-mappability

One of the requirements for some applications may be that the data recorded in particular fields has to be mapped to a particular classification or grouping scheme.

One way to simplify this process is for the application to check mappability at the time of data entry. If a selected Concept has no unambiguous map, the application may encourage or compel the user to refine their selection until a mappable Concept has been selected.

This type of facility should not be applied in situations where it may inappropriately affect the perceived accuracy or detail of a clinical record.

Constraining data entry based on context

All fields (data elements) used for data entry must be analyzed to understand what underlying context is implied. The appropriate concepts should then be selected for the value set of each field. Concepts from the Clinical Findings, Procedures, and Observable entities hierarchies can be used directly if the soft default assumptions are true. Otherwise, concepts from the Concept-Dependent hierarchy should be selected.

Particular care must be taken with systems that enable post-coordinated constructs to ensure that the appropriate context attributes are included.

To pre-coordinate concepts that do not already exist in SNOMED CT, care must also be taken to determine if any axis modification is shifting the meaning of a concept so it should move to the Situation with explicit context hierarchy.

Absolute and configurable constraints

Some of the constraints on data entry discussed in the preceding sections are absolute while others should be configurable.

- An application should not allow Inactive Concepts or any Special Concepts to be recorded.

Constraints based on subtype hierarchies, Subsets and Cross Maps should usually be configurable to particular institutions, users and/or data entry fields.

7.2.7 Configuring and applying data entry constraints

Overview

The previous sections describe various mechanisms for extending and constraining search and navigation during data entry. The scope of applicability of these facilities varies and these variations affect the way in which they may be implemented.

A few constraints apply to all data entry events in a particular application. These fixed constraints could be hard-coded in the application or explicitly optimized when importing and indexing SNOMED CT content.

Example:

One example is to exclude Inactive Concepts and Descriptions from searches. Before building this type of facility into an application, care should be taken to consider circumstances, such as creation and editing of queries where access to Inactive Concepts may be required.

Most search constraints are to some extent configurable and these require greater flexibility in the application design. There are several types of configurability that may be required. These range from installation configuration to context-specific dynamic configuration.

Installation configuration of data entry

Requirements of an organization that are general to all users may be applied when installing the application or when importing or indexing SNOMED CT content. These may include:

- Language Subsets, which constrain searches to the local language and dialect.
- Realm Concept Subsets and Realm Description Subsets, which apply national or organization constraints applicable to all users of the application.
- Realm Concept Subsets and Realm Description Subsets, which apply constraints applicable to all the clinical disciplines or specialties that use the installed system. For example, installations that are not intended for use in veterinary medicine will apply Subsets that exclude specific veterinary Concepts and Descriptions.
- A Navigation Subset that provides a data entry hierarchy appropriate to the needs of all users within an organization.

Log-on configuration of data entry

The application should allow search constraints that are specific to a particular user or group of users when loading or logging on to an application. The range of possible search constraints may be preset at installation but it should be possible to apply the user profile constraints without a significant delay. Uses of this type of configuration include:

- Realm Concept Subsets and Realm Description Subsets, which apply constraints or optimizations applicable to a particular specialty.
- Navigation Subsets that provide a restricted or extended data entry hierarchy appropriate to the needs or preferences of a particular specialty or user.
- Language Subsets that meet the needs of particular users in a multi-lingual environment.

Consideration should be given to requirements for this type of search configuration to be modified by a user or system administrator.

Dynamic reconfiguration of data entry

Constraints that assist fast and consistent routine data entry may sometimes need to be relaxed to enable more complex entries to be made.

- If a Navigation Subset limits the scope of hierarchical navigation, the application should enable the user to utilize the subtype hierarchy to allow other options or a more complete set of options to be reviewed.
- If a user is unable to locate the Concept that they require, it may be useful to enable some or all of the search constraints to be temporarily lifted.

Context-sensitive of data entry constraints

Some constraints may apply to particular data entry contexts. To support this type of functionality, an application should be able to switch between sets of search constraints in real-time. The constraints need to change instantly as a user moves between different data entry fields. Context-dependent constraints may include:

- Limitation of a search to subtype ancestors of an appropriate Concept.

Example:

A field for entry of a procedure may be associated with a constraint that limits searches to subtypes of the Concept "procedure."

- Limitation of a search to the Concepts or Descriptions that are members of an appropriate Context Concept Subset or Context Description Subset.

Example:

A field for entry of a laboratory service request may constrain searches to a list of valid investigations supported by a particular laboratory.

- Use of a particular Navigation Subset or a specified sub-branch of a Navigation Subset.
  - This is an alternative approach that may be used to allow more sophisticated control of data entry in a particular context.

7.2.8 Entering qualifiers and other post-coordinated representations

Overview

SNOMED CT contains many pre-coordinated Concepts that allow fairly complex Concepts to be represented by a single ConceptID. It also permits the qualification or refinement of Concepts to represent more detailed Concepts by post-coordinated combinations of several ConceptIDs.

Several types of post-coordinated data are outlined in this section from the perspective of data entry. These include refinement, qualification and combination. The requirements for and relevance of each of these will depend on decisions about data representation within patient records.

A syntax for representing post-coordinated expressions has been developed and is awaiting approval.

Entering refined defining characteristics

The application may allow a user to refine a Concept by selecting a subtype of one of its defining characteristics.

Example:

One of the defining characteristics the Concept "total replacement of hip" is "using"="hip prosthesis." The Concept "total replacement of hip" could be refined by allowing the user to specify one of the subtypes of "hip prosthesis."

Refinement options may be entered by selecting from hierarchical lists showing subtype values for each of the refinable characteristics. Simple lists or option buttons could support selection from limited sets of possible refinements. Wider ranges of potential refinement could be facilitated by text searches constrained to subtypes of one or more of the refinable characteristics.

Refinement should not be allowed for defining characteristics with the refinability value "not refinable."

Caution!

Some concepts should not be refined if the result means the new concept is not a subtype of the parent concept.

This situation occurs when context such as "Family history" or "Planned Procedure" is attached to a Clinical Finding or Procedure. "Family history" of a Clinical Finding needs to be defined in the situation with explicit context hierarchy. All post-coordinated constructs should consider the impact of context.

Entering sanctioned qualifiers

The application should allow a user to qualify a Concept by selecting one of its qualifying characteristics. These are represented by Relationships with a CharacteristicType value that indicates that they are optional qualifications rather than defining characteristics.

Entry of qualifiers may be supported by allowing selection from simple lists or by appropriate sets of option buttons. Automatic selection of qualifiers by parsing a search string may reduce the need for direct entry of some qualifiers.

The application should support refinement of qualifying characteristics in accordance with the refinability field value. The refinability value "mandatory to refine" indicates that entering the unrefined qualifier adds no useful information. In these cases, the application should force refinement of the selected qualifying characteristic to one of the subtypes of the qualifier value.

Entry of unsanctioned qualifiers

The application may also permit refinement of a Concept by the addition of qualifiers that are not sanctioned by inclusion as qualifying characteristics in the Relationships Table. These qualifiers may be constructed using pairs of Concepts, a subtype of the top-level Concept "attribute" combined with an appropriate value.

Any facility to allow qualification of Concepts in this way carries a risk of creating nonsensical or contradictory statements. It may also result in incomplete or inappropriate retrieval where the qualifier significantly affects the meaning of the Concept.

Constraints on the entry of qualifiers

Qualifiers should only be used where the result of applying them results in a true subtype of the original Concept. Therefore, qualifiers should **not** be used for the following purposes:

- Negation

Example:

"Fracture of humerus" should not be qualified by "excluded."

It would be inappropriate for data retrieval to treat this as a subtype of the clinical finding "Fracture of humerus."

- Certainty

Example:

"Carcinoma of cervix" should not be qualified by "possible."

It would be inappropriate for data retrieval to treat this as a subtype of the diagnosis of "Carcinoma of cervix."

- Subject of information

Example:

"Diabetes mellitus" should not be qualified by "family history."

It would be inappropriate for data retrieval to treat this as a subtype of the diagnosis of "Diabetes mellitus" in the patient.

- Planning stage

Example:

"Hip replacement" should not be qualified by "planned" or "requested."

A count of "Hip replacement" operations performed should not include this. Decision support protocols should not assume the patient has had this operation.

These and similar major modifications need to be handled in ways that are explicit and ensure that queries and decision support protocols are able to accurately retrieve and analyze the available information.

Entry of concepts combinations

The application may allow other combinations of Concepts in a single statement where a Concept that represents the full scope of an activity is not available. This approach might for example be applied where a single procedure, which lacks a pre-coordinated SNOMED CT representation, is a combination of two procedures that can be separately represented in SNOMED CT.

Facilities for entering combined Concepts should be implemented and used with care. It is appropriate to use these facilities when the combined result is conceived as a single statement that could potentially be used in many different patient records.

Example:

A diagnosis of "*gallstones with cholecystitis*" could be entered by selecting the "*gallstone (disorder)*" 235919008 and then selecting "*cholecystitis (disorder)*" 76581006 and combining these in a single statement[11].

It is not appropriate to use these constructs to attempt to express an entire encounter, episode or clinical history in a single statement.

[11] There is also a pre-coordinated Concept "*Calculus of gallbladder with cholecystitis*" which is equivalent to this post-coordinated combination.

Example:

If a patient is treated for "*gallstones with cholecystitis*" diagnosed by "*ultrasonography of biliary trac*t" with a course of "amoxycillin" followed after the acute phase has resolved by a "cholecystectomy," this should **not** be entered as a single complex post-coordinated statement combining the diagnosis, investigation and treatments.

7.2.9 Template and protocol driven data entry

In many healthcare disciplines similar data sets are collected for each patient. Clinical consultations for many conditions involve repeatable sequences of data entry. These structured and predictable data entry requirements can be met using sets of customized data entry fields or forms (templates) designed to collect particular data items. These data entry templates may be presented in a predefined sequence, as selected by the user. Alternatively the sequence of data entry may follow a branching pathway with previously entered data determining which branches are taken (protocols).

When using a structured data entry mechanism, SNOMED CT encoded data can be selected in a variety of ways. Some of these involve direct selection of Concepts and Descriptions while in others the encoding may result from responses to simple choices or entry of particular data values. The following list outlines some of the possible mechanisms for SNOMED CT encoding during structured data entry:

- User selection from a small list of possible Descriptions applicable to a particular field in a template or step in a protocol.
    - A Context Description Subset may specify the set of applicable Descriptions.
- Text search limited to a set of Concepts applicable to a particular field in a template or step in a protocol.
    - A Context Concept Subset may specify the set of applicable Concepts.
    - Alternatively the applicable Concepts may be specified as the subtype descendants of a single Concept.
- Association of a Concept with particular options presented by a check box, option button or other data entry control.
    - When selections are made using this control the appropriate ConceptID is added to the record.
- Association of a Concept with a data control used for entering a numeric or other value.
    - When a value is entered in this control it is labeled with the appropriated ConceptID.
- Association of a Concept with a particular combination of values or the result of a computation involving several items of previously entered data.
    - In its simplest form this is an extension of one or both of the previous options.
    - In some applications, information derived from the user-entered data, by decision support tools, may be encoded in this way.

Some installations allow free text to be entered at point of care if a needed concept is not included in the predefined short list. This text is then reviewed by trained staff who can then search and find the appropriate Concept, request the addition of a new Concept, and/or request that the Concept be added to the template's short list for future use. The success of this option relies upon trained staff who are available to do the review on a timely schedule, and the willingness of the clinician to use this approach sparingly, as it is greatly preferred to choose the appropriate concept and not enter free text.

7.3 Storing SNOMED CT Information in Records

7.3.1 Pre-coordinated and post-coordinated representations

Pre-coordination

The simplest form in which any concept can be stored is as a single identifier. This is referred to as pre-coordination, because all aspects of a potentially multifaceted concept are pre-coordinated into a single discreet form.

SNOMED CT contains more than a quarter of a million Concepts, and thus allows a wide range of clinical statements to be expressed in pre-coordinated form.

Example:

"*Laparoscopic emergency appendectomy*" has at least three distinct facets: "removal of appendix," "using a laparoscope," "as an emergency procedure."

SNOMED CT includes a Concept that pre-coordinates these facets. This allows "*laparoscopic emergency appendectomy*" to be expressed as a single ConceptID

Post-coordination

A multi-faceted concept can be stored using a combination of identifiers for its individual facets. This is referred to as post-coordination, because the various aspects of the concept are coordinated during data entry rather than in the preparation of the terminology. Three types of post-coordination are described in the following sections.

Post-coordination by qualification

Qualification is a type of post-coordination in which a concept is qualified by appending one or more qualifying characteristics.

Example:

"*Laparoscopic emergency appendectomy*" could be represented using identifiers for the concept "appendectomy" with identifiers for the following qualifiers: "using a laparoscope" and "as an emergency procedure."

SNOMED CT includes Concepts that act as attributes and values allowing this post-coordinated form to be constructed using the ConceptIDs for the following five Concepts:

- "appendectomy"
- + attribute="using" value= "laparoscope"
- + attribute="priority" values="emergency"

Post-coordination by refinement

Refinement is a type of post-coordination which is very similar to qualification. Rather than adding a qualifier, refinement specifies a subtype of a defining characteristic associated with a concept.

Example:

"*Total replacement of hip using a Sheehan total hip prosthesis*" could be represented by the Concept "*total replacement of hip*" with its defining characteristic "using hip prosthesis" refined to specify "*using a Sheehan total hip prosthesis*."

In SNOMED CT the distinction between refinement and qualification is dependent on which attributes are regarded as defining. The stored representations, qualifiers and refinements are identical. Therefore, this post-coordinated form can be constructed using the ConceptIDs for the following three Concepts:

- "total replacement of hip"
- + attribute="using" value="Sheehan total hip prosthesis"

Other important examples are in the Situation with explicit context hierarchy.

Example:

"*Family History of Longevity*" is a refinement of the category "Family History of" and the subtype "Family History Finding."

Post-coordination by combination

Example:

"*Gallstones with cholecystitis*" could be represented by combining the concepts for the disorders "gallstones" and "cholecystitis" as a single post-coordinated statement. Neither of these concepts is really a qualifier of the other since it could equally well be regarded as "cholecystitis will gallstones."

SNOMED CT allows Concepts to be combined in post-coordinated statement.

Combinations like this should only be used to represent concepts that can be regarded as discreet reusable clinical statements. They should not be used to construct arbitrarily complex representations of multiple statements to a particular record.

Some concepts, such as the first and last examples above, can be represented in either a post-coordinated or pre-coordinated form. However, there are other concepts, like the second example above, for which no pre-coordinated Concept exists in SNOMED CT. Although future releases of SNOMED CT will include new pre-coordinated Concepts, there will always be some clinical Concepts that require post-coordination.

Representing post-coordination

This guide does not specify a single right way to represent post-coordinated SNOMED CT information. Alternative representations have different profiles of advantages and disadvantages. The choice of representation depends on functional requirements in terms of flexibility, performance and the ability of the terminology representation to be used in existing information models (data schema) of software applications and/or to be used for communication between systems (e.g. the representations specific for particular messages or interfaces).

Some alternative representations are summarized below. These summaries are illustrations of the main options and do not go into extensive technical detail. More detailed design of each of these will lead to further alternatives that are not documented here.

Unrestricted relational representation

An unrestricted relational representation of post-coordinated SNOMED CT information is summarized below:

- Each clinical statement is recorded as a row in a relational table.
- The clinical statements table contains a field for a ConceptID.
- The clinical statements table has a one-to-many join to a qualifier table.
- Each row in the qualifier table contains a pair of ConceptIDs representing an attribute value pair (which refines or qualifies the associated Concept in the clinical statement table) and an optional Relationship group number to link associated attributes.
- Combined Concepts may be represented by explicitly combining two rows of the clinical statements table.

Parsable text representation

A way to represent post-coordinated SNOMED CT information as a simple parsable text string is summarized below:

- Each clinical statement is recorded as a row in a relational table or as an element in an XML representation.
- The clinical statements table (or element) contains a field (or element) for representation of the concept.
- The concept field (or element) contains a text string that encapsulates a post-coordinated representation of the concept according to a parsable syntax specified for this purpose. A possible syntax is suggested in Table 7.1.

Table 7.1: Possible textual representation of post-coordinated concepts

Key to symbols used

Symbols	Meaning	
+	Combines the ConceptIDs to the left and right of the symbol.	
=	Applies the value specified by the ConceptID to the right of the symbol to the attribute specified by the ConceptID to the left of the symbol.	
(	Starts an attribute value pair.	
)	Ends an attribute value pair.	
{	Starts a Relationship group	These symbols are only required where there is a requirement to distinguish multiple groups of qualifiers or refinements.
}	Ends a Relationship group	

Concepts used in examples

ConceptID	Meaning	ConceptID	Meaning
19227008	Foreign body	235919008	Gallstones
52734007	Total replacement of hip	260686004	Method
69695003	Stomach structure	261583007	Using
76581006	Cholecystitis	314580008	Sheehan total hip prosthesis
118828004	Procedure on gastrointestinal tract	363700003	Direct morphology
129287005	Incision – action	363704007	Procedure site
129303008	Removal – action		

Examples of the post-coordinated representation

Representation	Meaning
52734007(261583007=314580008)	Total replacement of hip using Sheehan total hip prosthesis
235919008+76581006	Gallstones with cholecystitis
118828004{(260686004=129287005)(63704007=69695003)} {(260686004=129303008)(363700003=19227008)}	Procedure on gastrointestinal tract with incision of the stomach and removal of a foreign body.

XML Representations

A way to represent post-coordinated SNOMED CT information as an XML element is summarized below:

- Each clinical statement is recorded as a row in a relational table or as an element in an XML representation.
- The clinical statements table (or element) contains a field (or element) for representation of the concept.

- The concept field (or element) contains an XML expression that encapsulates a post-coordinated representation of the concept according to a parsable syntax specified for this purpose.
  - Various alternative XML representations could fulfill this role. One option is the XML representation of the HL7 Version 3 concept-descriptor data type (CD). This is a topic for the IHTSDO.

Restricted relational representation

An alternative restricted relational representation of post-coordinated SNOMED CT information is summarized below:

- Each clinical statement is recorded as a row in a relational table.
- The clinical statements table contains a field for a ConceptID.
- The clinical statements table also contains fields for a specified number of qualifiers. These fields may be provided in different ways:
  - Each qualifier is represented by two ConceptID fields (one for the attribute and one for the value) and an optional field for Relationship group field. With this option the only restriction is the total number of qualifiers or modifiers that can be stored for each Concept.
  - Each qualifier is represented as a single ConceptID and carries the value of a qualifier attribute specific to that field. This restricts the usable qualifiers to those specified in the database schema.
  - Similar to above, but with different sets of qualifying attributes available according to the semantic type of the primary Concept in the statement. There are various ways of implementing this approach to ensure that the appropriate interpretation is applied to each row of the table.
- Combined Concepts may be represented by explicitly combining two rows of the clinical statements table.

Unlike the representations discussed in previous subsections, this approach limits the expressivity of post-coordinated statements. The advantage of this restricted approach is that it reduces the number of joins involved in retrieval queries. In some software environments this may significantly improve performance.

The balance between demands for flexibility and performance depends on user requirements. Therefore, limitations in expressivity may be acceptable for some users or user communities but not for others. However, it should be noted that these limitations might cause difficulties when communications are received from systems that support richer forms of expression.

Equivalent pre- and post-coordinated representations

The theoretical range of equivalent representational forms for a concept includes two end-points fully pre-coordinated and fully post-coordinated. In between these end points are a variable number of equivalent partial post-coordinated representations.

To illustrate this, Table 7.2 defines the hypothetical[12] concept "red aluminum pedal bike" and its supertype ancestors. Based on these definitions, Figure 7.1 illustrates a range of forms that could be used to represent the concept "red aluminum pedal bike." Form A is the most fully

[12] This hypothetical concept is chosen in preference to a real SNOMED CT concept to allow illustration of theoretical points with simple qualifiers. While all the points illustrated apply to some SNOMED CT concepts but there is no single concept that readily illustrates all these points.

post-coordinated representation while form K is fully pre-coordinated. Each of the intermediate forms are equivalent and assuming the accuracy of the definitions, each of these can be transformed by following subtype Relationships and appropriately adjusting the associated qualifiers.

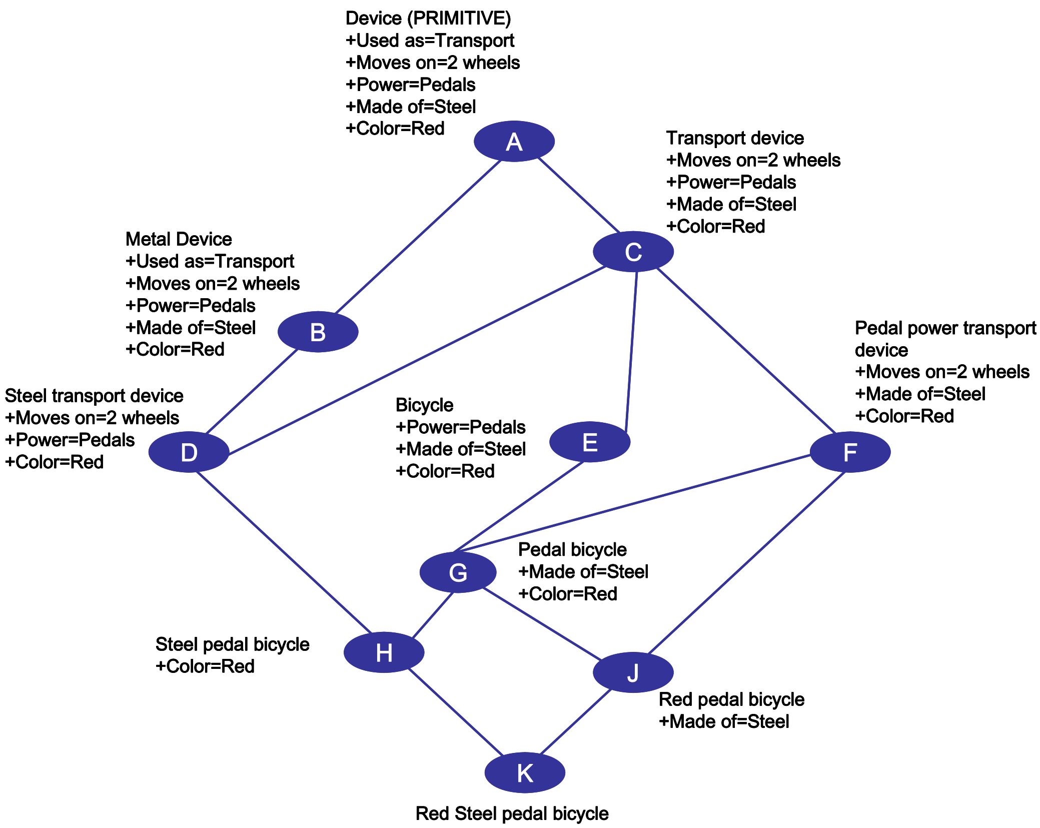


Figure 7.1: Illustration of alternative representations of "red steel pedal bicycle"

Table 7.2: Definitions of Concepts used in illustration of alternative representations

<table>
<tr><th>Concept</th><th colspan="6">Defining Characteristics</th></tr>
<tr><td rowspan="3">Red steel pedal bike</td><td colspan="6">IS_A = Red pedal bicycle
IS_A = Steel pedal bicycle</td></tr>
<tr><td>Used as</td><td>Moves on</td><td>Power</td><td>Made of</td><td>Color</td><td>Origin</td></tr>
<tr><td>Transport</td><td>2 wheels</td><td>Pedals</td><td>Steel</td><td>Red</td><td>Man made</td></tr>
<tr><td rowspan="3">Steel pedal bicycle</td><td colspan="6">IS_A = Pedal bicycle
IS_A = Steel transport device</td></tr>
<tr><td>Used as</td><td>Moves on</td><td>Power</td><td>Made of</td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td>2 wheels</td><td>Pedals</td><td>Steel</td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Red pedal bicycle</td><td colspan="6">IS_A = Pedal bicycle
IS_A = Pedal powered transport device</td></tr>
<tr><td>Used as</td><td>Moves on</td><td>Power</td><td></td><td>Color</td><td>Origin</td></tr>
<tr><td>Transport</td><td>2 wheels</td><td>Pedals</td><td></td><td>Red</td><td>Man made</td></tr>
<tr><td rowspan="3">Pedal bicycle</td><td colspan="6">IS_A = Pedal powered transport device
IS_A = Bicycle</td></tr>
<tr><td>Used as</td><td>Moves on</td><td>Power</td><td></td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td>2 wheels</td><td>Pedals</td><td></td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Bicycle</td><td colspan="6">IS_A = Transport device</td></tr>
<tr><td>Used as</td><td>Moves on</td><td></td><td></td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td>2 wheels</td><td></td><td></td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Pedal powered transport device</td><td colspan="6">IS_A = Transport device</td></tr>
<tr><td>Used as</td><td></td><td>Power</td><td></td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td></td><td>Pedals</td><td></td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Steel transport device</td><td colspan="6">IS_A = Transport device
IS_A = Metal device</td></tr>
<tr><td>Used as</td><td></td><td></td><td>Made of</td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td></td><td></td><td>Steel</td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Transport device</td><td colspan="6">IS_A = Device</td></tr>
<tr><td>Used as</td><td></td><td></td><td></td><td></td><td>Origin</td></tr>
<tr><td>Transport</td><td></td><td></td><td></td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Metal device</td><td colspan="6">IS_A = Device</td></tr>
<tr><td></td><td></td><td></td><td>Made of</td><td></td><td>Origin</td></tr>
<tr><td></td><td></td><td></td><td>Metal</td><td></td><td>Man made</td></tr>
<tr><td rowspan="3">Device (PRIMITIVE)</td><td colspan="6">IS_A = Thing</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Origin</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Man made</td></tr>
</table>

Two rules limit the range of equivalent representations:

Rule 1: It is not possible to transform a primitive Concept into a post-coordinated form.

- The IsPrimitive field indicates whether a Concept is Primitive.
- A primitive Concept has a facet that is not represented by its defining characteristics and therefore any attempt to represent it in a post-coordinated form results in a loss of information.

This is illustrated by consideration of the definitions of the concept “bicycle” in Table 7.2. The definition stated in the table is as follows:

“IS_A” = “Transport device”

+ “Used as” = “Transport” + “Moves on” = “2wheels” + “Origin” = “Man made”

This definition would also apply to a horse-drawn cart or a trailer. Therefore the concept “bicycle” must be regarded as primitive. Recognizing this fact means that some of the apparently equivalent representations in Figure 7.1 are incorrect. To correct this error any representation that is not a subtype of “bicycle” must explicitly state “IS_A” = “Bicycle” as shown in Figure 7.2.

Examining these definitions, it is apparent that the characteristics shown in gray are redundant because they are part of the definition of “bicycle.”

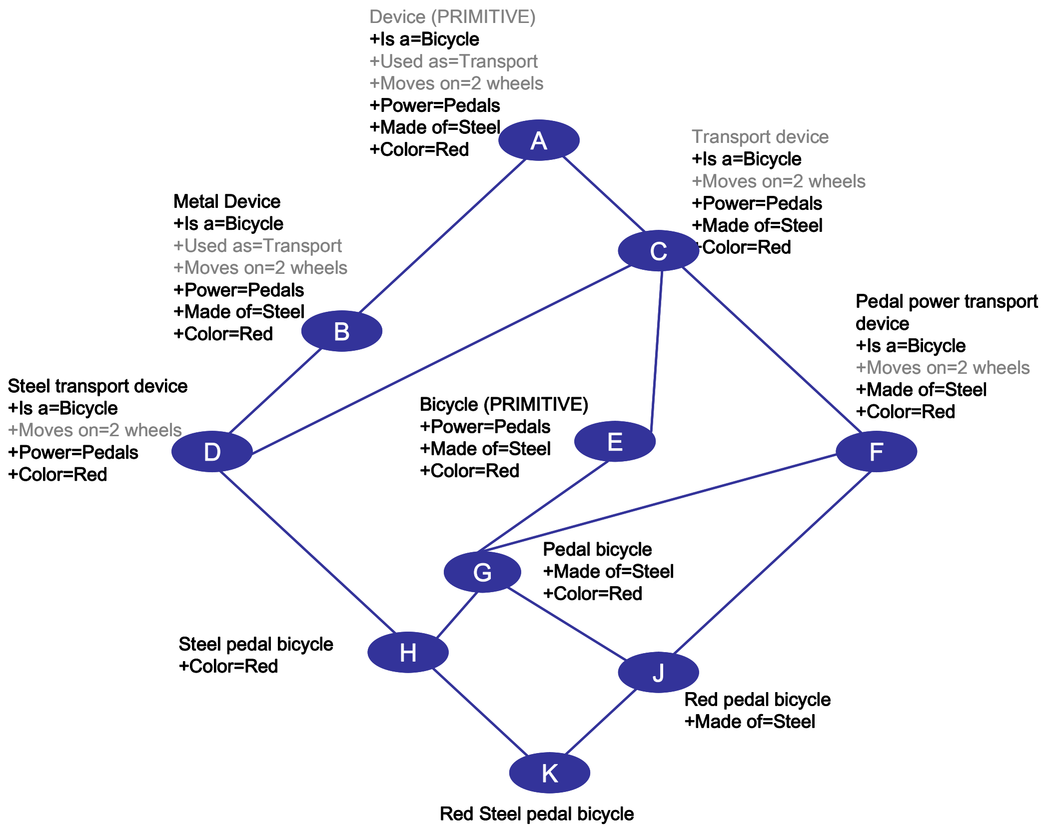


Figure 7.2: Representations of “red steel pedal bicycle” with “bicycle” recognized as primitive

Rule 2: It is not possible to transform a post-coordinated Concept into a fully pre-coordinated Concept unless such a Concept already exists in the released terminology.

This second rule is perhaps self-evident but it is stated because, like the first rule, it alters the available representations. If the concept "red steel pedal bicycle" was not available in a pre-coordinated form, there are two distinct representations that are as pre-coordinated as possible (i.e. "steel pedal bicycle" + "color" = "red" and "red pedal bicycle" + "made of" = "steel"). This is illustrated in Figure 7.3. In such cases there is no obvious reason to prefer one compared to the other.

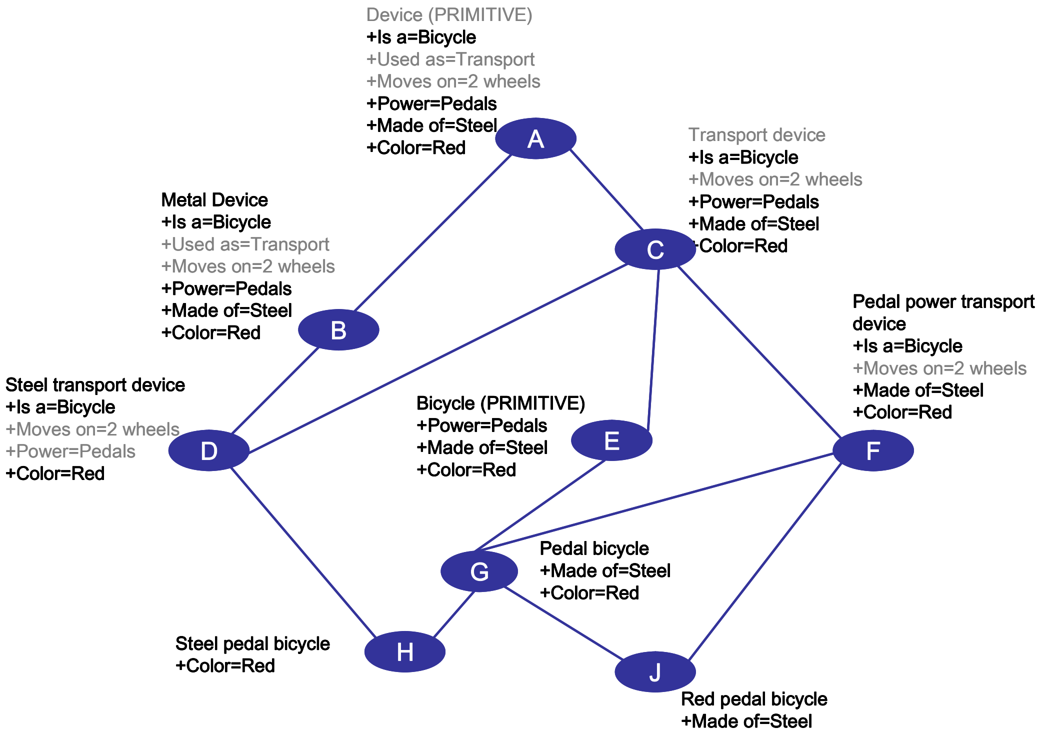


Figure 7.3: Representations of "red steel pedal bicycle" without pre-coordinated option

Transformation between pre- and post-coordinated representations

The previous section illustrates the possibilities for transformation between various degrees of pre- and post-coordinated representations. Transforming a Concept to a single form of representation is likely to enhance the efficiency of data retrieval. However, the previous section also shows that apparently minor corrections to the definition of a Concept can radically alter the set of equivalent representations.

Therefore, it is recommended that:

- The primary or original record should be stored using the representation that is as close as possible to the form in which it was recorded.
- If transformations to alternative representations are used to enhance the efficiency of retrieval, these should be stored as secondary supporting tables or indices.
  - This has the advantage that these alternative forms can be regenerated based on the most up to date set of definitions when a new release of SNOMED CT is installed, without affecting the integrity of the original records.

For further consideration of the form s of representation that may usefully assist retrieval, see [7.3.3].

7.3.2 SNOMED CT storage issues for electronic health records

Storing Concepts in electronic health records

Information in an electronichealth record should accurately reflect the way it was recorded by its author. If the author of a statement in the clinical record chooses a particular form of representation the system should faithfully store the information in that form.

Following this principle, the recommended approach for representation of SNOMED CT in a electronic health records is as follows:

- If, during data entry, an author selects a single pre-coordinated Concept to represent a clinical statement, the identifier of that Concept should be stored in the record.
  - This form of representation should remain as the original record of that statement. It should not be replaced by an apparently equivalent post-coordinated transformation of this Concept.
- If, during data entry, an author constructs a clinical statement by selecting a Concept and one ore more qualifier values, refinements or additional Concepts, the identifier of all the relevant Concepts should be stored in the record in a manner that reflects the relationships between them.
  - This form of representation should remain as the original record of that statement. It should not be replaced by an apparently equivalent transformation of this Concept into a pre-coordinated or differently constructed post-coordinated form.

An application should prompt for author endorsement of any alternative form of representation that it proposes to store in the original electronic health record. In this case, if the author accepts the alternative form presented by the application, this form should be stored as the original record.

The forms in which a technical implementer may wish to store data for efficient retrieval may differ from the forms dictated by the principles appropriate to storage of original entries in a electronic health record. However, it is recommended that any retrieval-oriented representation should be derived from rather than replace the original form of the record.

Storing terms

A electronic health record should also store the terms that were actually displayed to and selected by the author of the record. In some Realms the DescriptionID may be regarded as an adequate proxy for the full representation of the associated Term. However, in other jurisdictions there may be a requirement to store the original text as entered or selected by the user.

Storing the DescriptionID has the added advantage if a Description is found to be wrongly associated with a particular Concept or if the associated Concept is found to have non-synonymous Descriptions. In these cases, the DescriptionID can be used to map the information to the appropriate disambiguated Concept.

Maintaining integrity following SNOMED CT releases

The original recorded form of each clinical statement should be retained even when a new release of SNOMED CT changes the status of a Concept already stored in the record.

SNOMED CT releases contain Inactive Concepts and Descriptions to support the retention of original records in the form in which they were entered.

SNOMED CT Component History, Reference and Relationship tables contain information that allows data originally recorded using Inactive Concepts to be appropriately mapped to Active Concepts. However, updates based on this information should be appended to rather than replace the original representation. This ensures that if a subsequent release includes an improved mapping of the original Concept this can be utilized to enhance data quality.

7.3.3 SNOMED CT storage options for effective retrieval

Overview

The form in which records are represented may have a substantial impact on the efficiency, accuracy and completeness of retrieval. The forms that best suit retrieval may differ from the forms that are required to meet the principles of clinically safe and legally valid electronic health records.

Storing information as entered

This option leaves information in the form entered in the electronic health record with no additions to assist future retrieval. The application must do all the work needed to locate the required records and compute subsumption and equivalence when a request is made to retrieve data.

Minimizing post-coordination

One possible approach to optimization of retrieval is to transform the original stored information into an equivalent representation with the minimum number of post-coordinated components.

The objective of this approach is to allow the generation of simple indices for the pre-coordinated representation. It is then possible to undertake most retrievals using the "is a" subtype hierarchy to compute whether Concepts in the record are subtypes of the Concepts used to specify retrieval. Where post-coordination is required, the minimum number of additional tests are required to confirm that a Concept in the record meets the specified retrieval criteria.

One difficulty with this approach is that there may be more than one representation that requires the same degree of post-coordination. This is discussed in more detail and illustrated in [7.3.1].

If this approach is adopted additional rules need to be applied to determine the choice between alternatives with a similar number of post-coordinated components.

> **Example:**
>
> In the hypothetical example illustrated in Figure 7.3, the Concept "red steel pedal bicycle", for which no pre-coordinated representation exists, could be represented as:
>
> "red pedal bicycle" + "make of" = "steel"
>
> *or*
>
> "steel pedal bicycle" + "color" = "red"

Both are equally close to the objective of minimizing post-coordination. A rule is needed to determine which of these is preferred. There is no obvious right or wrong solution to this but a simple rule that places the attributes in an order will, if applied consistently, allow all post-coordinated representations to be reduced to a single minimized form.

Maximizing post-coordination

An alternative approach is to expand any pre-coordinated concepts in the record to their fullest possible post-coordinated forms. This general type of transformation is illustrated in [7.3.1].

This approach requires a richer record structure but has the advantage that there are three possible end-points to post-coordination, each of which ensures that any computably equivalent representations of Concepts will expand to an identical post-coordinated form. The three end-points are summarized here:

- Short canonical form
  - This is the most parsimonious of the three options.
  - A concept is represented as the combination of:
    - Subtype relationships with its most proximate primitive supertypes.
    - The recorded qualifier values and/or defining characteristics that distinguish it from its most proximate primitive supertypes.
- Long canonical form
  - This option is more verbose as it includes some redundancy.
  - A concept is represented as the combination of:
    - Subtype relationships with its most proximate primitive supertypes.
    - All of its recorded qualifier values and/or defining characteristics, irrespective of whether they are shared by its most proximate primitive supertypes.
- Exhaustive post-coordinated form
  - This option is extremely verbose.
  - A Concept is represented as a combination of:
    - Subtype relationships with all of its supertype ancestors
    - All of its recorded qualifier values and/or defining characteristics, irrespective of whether they are shared by its most proximate primitive supertypes.

If the retrieval criteria are expressed in a similar form, a relatively simple query can interrogate the record for all entries with a matching set of primitive Concepts and specified characteristics.

A practical example of this approach is illustrated in D.3.3 in Appendix D.

7.3.4 Record architectures, structures and semantics

Record structure standards and proposals

The semantic model of SNOMED CT does not replace the need for a logically sound health record structure and the SNOMED organization does not specify a particular health record structure. SNOMED CT is a controlled terminology that can be used in many different health record systems. It is particularly well suited to use in conjunction with health records designed to meet standards including:

- The healthcare communication and structured document standards of HL7 (*www.hl7.org*)
- The European (CEN) Prestandard for Electronic Healthcare Record Communication (ENV13606) (*www.centc251.org*)

Current activities in HL7 will bring together the best aspects of this work to develop the next generation of global health record standards, in particular through the HL7 Electronic Health Record SIG, two initiatives that utilize HL7 Version 3 methodology to implement the principles of the ENV13606 and HL7 Structured Document TC.

SNOMED CT representations of clinical concepts meet the need of these structural and architectural standards for a consistent controlled coded terminology.

Using SNOMED CT in standard architectures

The broad principles of the established health record architectures are based on a layered structure of components that contain and provide context to lower level components.

The container structures include some or all of the following:

- A top-level component representing the entire health record of one person.
- Intermediate layers representing information from various sources.
- A fixed transaction/composition layer at which an entry or set of entries are attributed to (and possibly signed by) an author.
  - Examples of this level include consultation notes, letters, reports, and other documents.
- Further levels that represent logical grouping within a record covering:
  - Topics, heading and categories.
  - Cluster or batteries of closely associated information.

Within the containment structures are two lower level components:

- Clinical statements
  - A clinical statement may vary in structure to accommodate different kinds of information (e.g. patient history, clinical finding, investigation results, plans, procedures, medication and other therapies).
- Link statements
  - Link statements state associations between record components.
  - Links statements can be used to specify:
    - Problem-oriented groups of record components and viewing.
    - Causal and other specified links recorded by the author of a record entry.

Each health record component has the potential to include:

- Dates and times of actual and planned events
- Associations with people, organizations, devices and other entities that participate or are used in relations to a recorded event or plan.
- Codes or other representations that name or provide the semantic information container, link, or statement.
  - SNOMED CT fulfills this role in a structured health record.
- Additional data including text, numeric values, images and other digital data.

When SNOMED CT is used in a structured record, the links and temporal associations of components combined add further richness to the potential power of expression. This has significant advantages and is essential for many types of aggregation and decision support. However, it also adds a complicating factor that should be taken into account when designing, recording, storage, and retrieval facilities.

Example:

To retrieve and analyze the records of patients with two potentially related conditions such as "AIDS" and "Gastro-enteritis" it is not necessary for this combination to be represented in a single pre- or post-coordinated concept. Instead, it is possible to look for co-existence of the individual Concept "Gastro-enteritis" within the records of patients who also have "AIDS."

- The advantage of this is that there is no need for the clinician to have made the association between the two conditions. Therefore a more complete assessment of the incidence of "Gastro-enteritis" in patients with "AIDS" can be made.
- The disadvantage is that if a pre- or post-coordinated SNOMED CT representation of the combined concept is used, these records will not necessarily be computably equivalent to those with the two conditions recorded separately.

There is no absolute rule on when to use multiple statements associated using record structure constructs, and when to use intrinsic pre- or post-coordinated SNOMED CT representations. The decision maybe influences the functionality of a particular system and the specific user requirements that the system is serving. However, the following guideline is suggested:

- A combined pre- or post-coordinated representation is appropriate if:
  - The combined concept is a discrete recognizable Concept that differs in some way from the simple combination of the two concepts. For example:
    - "Diabetic cataract" is not the same as "Diabetes mellitus" + "Cataract" because other types of cataract may occur in the same patient.
    - "Fracture of radius and ulna" is a clinically recognizable injury, which is most effectively conveyed as a single concept.
- Separate records for each Concept are appropriate if any of the following apply:
  - The combined Concept represents the coincidence of two potentially associated conditions or procedures;
  - The temporal and other characteristics of the two Concepts are different;
  - Where the association between the two Concepts is causal.

Example:

"Fracture of femur" caused by "fall down stairs" should be represented as separated statements linked by an appropriate record structure component. The SNOMED CT Concept "Due to" could be used to name the link between these statements.

7.3.5 Safely representing the context of recorded codes

Introduction

A variety of contextual factors may affect the interpretation of statements. Contextual factors typically fall into the gray area between record structure and the semantic model. Some of these may have a profound impact on the meaning or interpretation of a statement.

This section divides this issue into four distinct categories:

- Contextual information that is not represented by SNOMED CT
- Structures that may be labeled using SNOMED CT
- Status terms that have a profound effect on SNOMED CT encoded statements
- Context that can safely be represented using qualifiers

Contextual information that is not represented by SNOMED CT

Clinical statements that contain SNOMED CT Concept representations will be associated with some information which is not intended to be represented using SNOMED CT:

- Dates, times of an activity of recording and activity
- Quantitative information including ranges and durations
- Identifiers or names of authors, providers of information or other parties involved in a recorded activity.

SNOMED CT is not intended to represent this information. Appropriate constructs in a standardized or proprietary record architecture should be used to relate this information to SNOMED CT encoded clinical statements.

Structures that may be labeled using SNOMED CT

Clinical statements may be contained within structures that represent collections of related information. According to the nature of these structures, SNOMED CT may be used to label them. However, care should be taken to ensure that any semantic implication from such a label is clearly specified.

Many labels (such as headings within a document) may be used only to organize information for a human reader. The existence of a label such as "plan" or "family history" (even if encoded using SNOMED CT) may not necessarily affect the computer interpretation of the data within it.

Implementers should take extreme care to ensure that any semantic implication that a human reader may assume from such labels is stored on the system in a manner that allows safe interpretation. It is recommended that any apparent inherited semantic context should be represented explicitly at the individual statement level.

Example:

If a data element "Family history" is used and the concept "diabetes mellitus" is encoded under that heading, the statement stored in the record should encapsulate the full semantics (i.e. Family history+(Associated finding=Diabetes mellitus) using the SNOMED CT Concept Model.

Other areas in which structures might be labeled with SNOMED CT Concepts include:

- Links between statements – SNOMED CT could be used to identify the nature of the link.

Example:

To indicate a presumed causal relationship.

- Indication of types (rather than identities) of people of organizations.

Example:

To indicate that the source of a piece of information was the patient themselves or a specified relative.

Context and Axis Modifiers

The following are examples of "axis modifiers" which may fundamentally alter the interpretation of information encoded using SNOMED CT:

- Subject of information

Example:

Stating that a relative of the patient has a particular disease. This may be recorded to state either "family history" or a "social context." If the disease is represented as a SNOMED CT Concept then it must be distinguishable from a statement that patient has that disease.

- Stage

Example:

Stating that a patient should have, has been referred for, or has declined to undergo a particular procedure. These must be distinguishable from statements that a patient has had the stated procedure.

- Negation and uncertainty

Example:

Stating that a diagnosis has been excluded or is unlikely.

- Contra-indication

Example:

Related to a treatment specified using its ConceptID.

There is a temptation to use these modifications as though they were qualifiers. This is not a safe practice because the assumption is that a qualifier refines the meaning of the qualified Concept. A refined Concept should always be a subtype of the original Concept. This is not the case for these major modifications as illustrated by the following:

- Records of a "family history of" + "diabetes" would not be expected in a response to a query for patients with a record of diabetes mellitus (and its subtypes).
- Records of "planned" + "hip replacement" should not be counted when analyzing the records of patients who have had any type of "joint replacement."
- Records that state "meningitis" + "excluded" should not be counted as cases of meningitis.
- Records that state the patient is "allergic to" + "penicillin" are not records of treatment with an antibiotic.

This issue is discussed in section [7.2.8] in relation to data entry.

The recommended approaches are:

- To ensure that the record structure captures this information in a consistent and reliable way that can be interpreted accurately when retrieving or communicating information.
- If post-coordinated SNOMED CT representations are used, the status Concept should be qualified by the condition or procedure, e.g.:
  - "family history" + ("associated finding" = "diabetes mellitus")

  is acceptable but **not**

  - "diabetes mellitus" + ("qualified by" = "family history")

Context that can safely be represented using qualifiers

Where a contextual modification can be logically regarded as a refinement of the original Concept it is reasonable to use a qualifier. Examples of this include "severity," "episodicity" and "laterality."

Concepts with built-in context

Some Concepts derived from earlier terminologies (i.e. SNOMED International and the Read Codes) contain built-in context. An example is the concept "FH: Diabetes mellitus" (FH being an abbreviation for family history). These concepts are in the Situation with explicit context hierarchy.

To the extent possible with released context attributes, these Concepts are defined (and will continue to be reviewed) so that they are computably equivalent to appropriate post-coordinated representations.

Example:

The concept "FH: Diabetes mellitus" is defined to be equivalent to the post-coordinated representation[13]

- "family history" + ("associated finding" = "diabetes mellitus")

[13] This is not the case in the first release as the appropriate defining characteristics are not in the release set.

7.4 Retrieving and Aggregating SNOMED CT Information

7.4.1 General aspect of selective retrieval

Requirements for selective retrieval

Selective retrieval is an essential function for a health record system. There are two main types of requirements:

- Retrieval of selected records from the records of members of a population of patients for one or more of the following purposes:
  - Aggregations and analysis of data to support:
    - Epidemiological monitoring
    - Clinical research
    - Audit of care delivery
    - Service planning
- Identification of patients with specific risk factors or other characteristics.
  - To allow specific preventative, investigative or therapeutic measures to be appropriately focused.
  - To allow further selective retrieval and analysis of the records of a subpopulation of patients.
  - To enable selection of patients for entry in a clinical trial.
- Retrieval of selected records from an individual patient record to enable:
  - Display of summary views and/or pre-completed template screens containing appropriate selected information.

Example:

Active problems/diagnoses, current medication, recent investigation results or blood pressure readings.

  - Automating responses to questions posed by a decision support protocol.

Example:

To check the record for specified symptoms, findings, investigations, procedures or diagnoses.

The specific features of different requirements are discussed in [7.4.4]. The following subsections address points that are common to all types of retrieval.

Retrieval performance

The following sections identify factors that may influence performance when undertaking selective data retrieval. There are no fixed rules for optimization of retrieval performance. Application developers should interpret the issues outlined in the guide in the light of their experience with the operating systems and data management tools that they use.

An evaluation of different approaches to retrieval was undertaken for the NHS, in connection with work on Clinical Terms Version 3 implementation. This showed that the "best" approach was not the same for all relational databases. Some software environments favor one approach while a different approach may be more effective in another environment. Therefore, it is likely that some of the factors discussed will have a significant impact on some developers, while being less relevant in others.

Retrieval quality

The quality of selective retrieval is measured in terms of two factors:

- **Completeness:** Retrieval should select all records that meet the selection criteria.
- **Specificity:** Retrieval should not select any records that do not meet the selection criteria.

The semantic structures of SNOMED CT assist application developers to achieve these goals by allowing alternative representations of similar information to be recognized.

Retrieval criteria involving record structure

Before addressing the specifics of SNOMED CT related retrieval criteria it is important to recognize that these only form one part of the picture. Most selective retrieval criteria will include a mixture of predicates, some of which apply to SNOMED CT encoded data and some of which apply to other data in the patient record. This non-SNOMED CT encoded data includes:

- Data directly related to coded clinical statements. This includes:
  - Dates and times (e.g. time of an event of finding).
  - Organizations, people or devices involved in a recorded activity or finding.
  - Temporal or causal relationships with other clinical activities or findings.
  - Quantitative values associated with SNOMED CT encoded statements.
  - Associated status and contextual information.
- Data related to the patient:
  - Age and sex;
  - Organizations and people responsible for care;
  - Occupation, pre-existing disorders or other known risk factors.

The interplay of these factors with SNOMED CT encoded data may affect the optimum approach for data retrieval. Some non-SNOMED CT encoded retrieval criteria may significantly reduce the potential set of patients or in patient record entries that qualify for retrieval. In these cases, it may be useful to apply these criteria before testing SNOMED CT specific criteria.

Example:

- A retrieval request for the rubella vaccination status of eight-year-old girls in a family practice with average population distribution requires the review of less than 1% of the population of records.
- A retrieval request for patients who have undergone a particular procedure in the last month only needs to review record entries made in the last month.
- A retrieval request for the most recent investigation results and current medication might be more processed by initially identifying a set of recent records. Checking these records for relevant SNOMED CT values may be more efficient than applying individual queries to the entire record for each of the required items of recent information.
- A retrieval request for people with a rare clinical condition, who also have a relatively common disorder, may be more efficient if the few people with the rare condition are selected first, limiting the scope of the query for the more common condition.

These examples illustrate a general point rather than to offer guidance on the specific searches. It is important to bear in mind that the performance, completeness and specificity of retrieval are dependent on the structure of the record as well as the semantics of SNOMED CT.

7.4.2 Retrieving records containing selected concepts and their subtypes

Introduction

Information in health records may be expressed at various levels of specificity.

Example:

To represent diagnoses of:

- Chest infection
- Left lower lobe pneumonia caused by pneumococcus

Criteria for selective retrieval may also need to be stated to different levels of detail.

Example:

To retrieve all records of

- Respiratory tract infections
- Left lower lobe pneumonia
- Pneumococcal pneumonia

Occasionally a query may be designed to retrieve only record entries that include a particular general Concept. This may be useful for a quality review or to find record entries that are too general to cross map to a required classification.

However, in most cases, a general query should include more specific Concepts recorded in the record. For example, if the selected Concept is "Respiratory tract infections" the user would expect record entries containing Concepts such as "Chest infection" or "Left lower lobe pneumonia caused by pneumococcus" to be retrieved. The subtype hierarchy of SNOMED CT is designed to facilitate this type of retrieval. Four techniques that can be used for this purpose are outlined in the following subsections.

Important Note:

The subtype hierarchy is improved with new releases of SNOMED CT. These changes need to be considered if more than one version of the hierarchies are used for data analysis.

Queries expanded to identify all subtypes

A query that explicitly includes the ConceptIDs of all subtype descendants of the Concept to be retrieved can be built using one of the following methods:

- A recursive tree-walk following "IS_A" Relationships – from the selection Concept to its subtypes and the subtypes of its subtype. Each branch of the tree walk ends on reaching a Concept with no subtypes or a Concept that is already in the set of selected Concepts.
- Using pre-generated branch number ranges associated with the selection Concept and looking up all Concepts with branch numbers in those ranges. This could be much faster than a tree-walk if Concepts are indexed by branch-number.

- Using a stored list of subtype ConceptIDs for frequently queried Concepts. This would initially be generated in one of the other methods and then reused in various queries. Any stored list would need to be rebuilt after installing each release of SNOMED CT.

The resulting query may contain a large list of potential ConceptIDs, but the actual query structure is simple. Therefore as long as the database engine does not restrict query size, this type of query can be run in any environment that support SQL or an SQL-like query language.

This technique is likely to be most effective when a large number of candidate record entries need to be examined and when Concept selection criteria are relatively narrow. Selecting all diagnoses using this approach would generate a predicate with tens of thousands of ConceptIDs. Extremely large queries may not perform efficiently or may fail to run in some environments.

Subtype tests on each recorded concept

The Concept recorded in each candidate record entry can be tested to determine whether it is a subtype of the Concept to be retrieved. The test can be applied in one of the following ways (see also Testing and traversing subtype relationships [6.2.5]):

- A recursive tree-walk following "is a" Relationships from the recorded Concept to its supertype and the supertypes of its supertypes. Each branch of the tree walk ends on reaching the Root Concept or a Concept that has already been visited. The test ends with a positive result if the selection Concept is encountered during the tree walk. Otherwise when all supertypes have been visited, the test ends with a negative result.
- Optimized subtype testing using techniques such as branch numbering and tree-walk enhanced with semantic-type identifiers or hierarchy flags.

This technique is likely to be effective when the number of candidate record entries to be examined is relatively small or if the Concept selection criteria are broad. Performance is directly dependent on the time taken for each subtype test. Therefore, extensive use of this approach may only be feasible by applying one or more of the optimizations suggested in section [6.2.5].

Use a database with built in hierarchical functionality

Some databases have features which build in hierarchical functionality. These databases may support extensions to SQL that allow a predicate to be specified in a way that implies that the database schema "understands" the subtype hierarchy.

Example:

It is possible to envision a statement such as:

WHERE Record.ConceptID SUBTYPE OF 8957000

If a database supports this type of predicate, it clearly simplifies the writing of SNOMED CT queries. It is also reasonable to assume that functionality of this type, built into a database engine rather than added as an afterthought, will deliver enhanced performance. However, this assumption should be tested as it depends on how appropriate the internal implementation is to subtype hierarchy of the size and complexity of SNOMED CT.

Branch-range indexing of individual records

The use of branch numbering has already been mentioned as a sub-option of two other approaches to subtype retrieval. The same branch numbers could be used to produce an index of all record entries stored in an application. The technique is as follows:

- Every record entry is indexed using the branch number of the Concept stored in that entry.
- The set of branch number ranges associated with the selection Concept is then used to query the branch number index.

This approach is likely to deliver high performance retrieval but it has a significant drawback. Branch numbers have to be regenerated after each SNOMED CT release and the numbering changes each time. Therefore, any indices based on branch numbers must also be rebuilt after each release, and until this rebuild is complete, this method cannot be used for retrieval. The previous set of branch numbers could be used for retrieval during the transition period but this requires a parallel set of branch numbers and branch number ranges.

The likelihood of enhanced retrieval performance should therefore be balanced against the addition of complexity to terminology updates and record maintenance.

Retrieval Based on other relationships

While many queries will use SNOMED CT's hierarchical subtypes to aggregate data, the attribute relationships can also be used. For example, to find all procedure concepts that use a laparoscope, search in the Relationships Table for Concepts with a relationship of Using Access Device: Laparoscope. Note that role hierarchies can be used to construct these queries.

7.4.3 Selective retrieval of post-coordinated concepts

Introduction

The previous section deals with the retrieval of records that contain pre-coordinated representations of Concepts. The mechanisms and methods discussed in that section need to be extended to cover post-coordinated representations.

The selective retrieval mechanisms applicable to post-coordinated information depend on the way in which this data is stored. If data is transformed to generate tables or indices that facilitate retrieval, the form of this derived data determines the type of mechanisms that can be used.

There are two significant factors in the completeness, specificity and performance:

- The structure used for representing post-coordination. See section [7.4.3.].
- Whether the information is only stored in the form entered of is also stored in a manner that seeks to enhance the consistency of post-coordination. See section [7.4.3].

Retrieval from unrestricted relational representations of post-coordinated data

Unrestricted relational representations provide a flexible medium for storage retrieval of post-coordinated data. A query can be specified at any level of detail to examine the primary Concept in a statement and any or all of the associated post-coordinated qualifications, modifications, or combinations.

However, the number of joins required to specify an appropriate query may affect performance.

- Each clinical statement consists of a row in one table joined to a row in a qualifier table for each post-coordinated addition. The clinical statement itself may have other structural relations (based on the record structure) and each patient record may consist of hundreds of thousands of statements.

The effect of this will vary according to the power and configuration of the relational database. However, some application developers may seek alternative, more limited representations to improve performance.

Retrieval from restricted post-coordinated information

An application may store data in a restricted relational representation, which limits post-coordination to a pre-specified set of qualifiers. Criteria for selection based on the values of a limited set of qualifiers require a minor extension to any of the approaches discussed in the previous section. However, there are two significant points to note:

- When applying criteria to the values of a qualifier, any subtype of the specified value should be selected. This is similar to the consideration for the primary Concept. However, the number of tests to be performed will be more limited because:
  - Typically a qualifier value will have relatively few subtypes.
  - Only record entries that match on other criteria need to be tested.
- Some of the supported qualifying attributes may also occur in defining characteristics of some Concepts. A query that specifies the presence of a particular qualifier must not miss these cases. One way to address this issue is to ensure that when storing or transforming data for retrieval, the value of any defining characteristics that are also supported, and qualifying attributes should be copied into the qualifying value field.

Retrieval from post-coordinated data stored as parsable text or XML

Parsable text strings or XML elements are not well suited to rapid retrieval from large populations of records. However, optimization is possible by augmenting the stored form with indexes to Concepts (e.g. indexing ConceptIDs or range number) or by using an XML-aware database. Without such optimization it may be possible to achieve acceptable performance for retrieval from individual records, documents or messages represented in a structured form using XML[14].

Retrieval of post-coordinated data stored as entered

Where post-coordinated data is only stored as entered, retrieval mechanism must do all the hard work of calculating the equivalence between statements expressed in different ways. This is possible for a small-scale search (e.g. within a single patient record) but across a large population of records it may be difficult to achieve an acceptable performance.

Retrieval from minimized post-coordinated forms

If post-coordination is deliberately minimized before storage, this allows most of the search process to be concerned with querying or testing subtype descendants.

If the query needs to specify selection criteria that cannot be expressed by a single Concept, further testing required. Even then, if there are rules for consistently minimizing post-coordination, most queries remain easy to construct and apply.

Some complex queries may present more difficulties with this approach but it remains a reasonable option for application developers concerned with minimizing the overhead related to storage and retrieval while delivering reasonable performance and flexibility.

[14] Examples of retrieval from XML representation and HL7 messages is provided in Appendix D.3.

Retrieval from maximized pre-coordinated forms

Deliberately maximization of post-coordination to one of the forms described in [7.3.3] offers them most flexible approach. Of the three forms suggested:

- The exhaustive form
  - Simplifies queries since everything that is true about a Concept is stated and there is no need to check subtype descents.
  - Carries a heavy storage penalty for every record stored.
  - Requires computation of the representation of each Concept after every release.
- The long canonical form (see also example in Appendix D [D.3.3])
  - Allows queries that are relatively simple provided that a mechanism exists for checking subtype descents.
  - Although more terse than the exhaustive approach, storing this information for every record stored still has a significant storage overhead.
  - Requires rechecking or re-computing after a release, but his can be done directly from the release files by combining the “IS_A” relationships in the Canonical Table with the other (i.e. not “IS_A”) defining characteristics in the Relationships Table.
- The short canonical form
  - Requires slightly more care in construction of queries than the long canonical form.
  - Requires slightly less storage than the other maximized forms.
  - Like the long canonical form, can be rebuilt directly from the release tables.

7.4.4 Requirements for specific uses of selective retrieval

Specifying retrieval requirements

An application should provide a mechanism to allow users to specify retrieval requirements using SNOMED CT. This facility should allow queries to be generated that combine SNOMED CT specific selection criteria with other health record criteria.

A terminology browser that combines text searches and subtype hierarchy navigation is likely to be essential for defining SNOMED CT specific selection criteria.

Facilities for testing and traversing subtype relationships may be need to expand selected Concepts into detailed queries that can be run against the stored records.

Additional facilities to apply combine Concepts or apply qualifiers or refinements may be necessary to support more sophisticated searches.

Selective retrieval for reporting and analysis

Population-based retrieval and reporting is usually a task that can be run in the background or scheduled for later execution. Therefore, real-time responses are usually not essential.

The process of analyzing a large number of records may take several minutes or perhaps even hours. If the application spends a little time generating an optimized query before starting to access the records, this is acceptable and may shorten overall execution time. Therefore, a technique such as query expansion may be appropriate for these tasks.

Users may also have requirements for reports on individual patients or a small group of patients. In some cases there may be an expectation of a real-time response to requests for these reports. If so, the delay while several selection criteria are expanded may be unacceptable. If the same criteria are used many times, storage of the expanded form may be a realistic option. Otherwise, an alternative retrieval technique should be considered.

Selective retrieval for decision support

Decision support tools are usually used during a consultation with a patient. Real-time response without significant delay is essential if these tools are to be used regularly and perceived as a help rather than a hindrance. A decision support algorithm may need to selectively retrieve several records to inform a single decision or piece of advice. Many of the retrieval criteria are likely to be quite general. The time taken to expand an apparently simple set of criteria so that they include all appropriate subtypes is likely to significantly impair performance. The expanded criteria could be stored in or associated with the protocol. However, the requirement to update these with new SNOMED CT releases and whenever the protocol changes add to the maintenance burden.

Since decision support protocols are primarily concerned with the records of an individual patient, it may be feasible to test all candidate records (e.g. all records that fall within a specified date range) to see if any of these are subtypes of the selection Concept(s).

Other alternatives should also be considered including those discussed in [7.4.2].

Decision support tools as authors of data in the record

As well as retrieving SNOMED CT encoded information a decision support tool may need to make entries in the record. These entries may arise directly from user interaction with a template or protocol. However, some entries made by decision support tools may record decisions made by or advice given by the tool.

Retrieving records encoded with Inactive Concepts

Records that have been encoded using Concepts that are no longer active can be retrieved by using the Historical Relationships (i.e. “same as,” “may be a,” “replaced by” and “was a”) in addition to “is a” subtype relationships.

An application should allow users to specify which (if any) of these Relationships should be followed when determining whether to retrieve a record entry.

- A sensible default is to treat duplicate Concepts related by “same as” Relationships and erroneous Concepts related by “replaced by” Relationships as though they were interchangeable with the related Concepts.
- In the case of the ambiguous Concepts related by “may be a” Relationships the solution is less clear-cut. A choice must be made between the importance of completeness, which is best served by including these Concepts, and selectivity, which is better served by excluding them.

Retrieving and analyzing legacy coded data

SNOMED CT can also be used for generated queries that examine legacy data recorded using SNOMED International, Clinical Terms Version 3 or earlier versions of the Read Codes. This can be done by using the approach outlined in [7.4.2] to generate a query that includes all the subtypes of a selection Concept. However, the appropriate legacy codes (i.e. The CTV3ID or SNOMEDID) are added to the query instead of the ConceptID.

7.5 Communicating SNOMED CT Information

7.5.1 Introduction

This section provides an outline of some general issues related to communication of SNOMED CT information using standard communication structures. Appendix D provides a more detailed example using a pre-ballot version of an HL7 Version 3 XML message.

7.5.2 Representation of SNOMED CT information in communications

Various media exist for communicating computer processable data between applications. These include:

- Messages
- Structured documents
- Portable storage media (smart cards, memory cards or other similar devices)
- Application interfaces (including COM / CORBA and programmable web interfaces).

A common feature of any method of communication is the need for a formal standard (or de-facto agreed) representation of the communicated information.

To enable full communication of SNOMED CT information these agreed standards must allow the communication of pre- or post-coordinated representations.

Some current standards do not provide explicit support for post-coordination. Examples include:

- HL7 Version 2.x messages
- EDIFACT implementations of European (CEN) Prestandards for laboratory communication used in by the UK NHS.

In these cases, it may still be possible to include post-coordinated information by agreeing to a syntactic representation that can be used in a single message element.

> **Example:**
>
> Subject to message field length constraints, the parsable text form illustrated in Table 7.1 could be included in place of a simple code.

The use of this type of technique is not recommended since it may distort the intended semantics of the message, but also, and more significantly, it requires the recipient to agree to parse the code in a particular way. There is no point sending a parsable text representation of a post-coordinated Concept to recipients with no understanding of that form of representation.

More recent standards make specific provision for the support of post-coordination in representations of clinical statements. Examples include:

- HL7 Version 3, which includes the “concept description” (CD) data type which provides unlimited scope for post-coordinated modifiers.
- European (CEN) Prestandard ENV13606 for Electronic Healthcare Record Communication, which include a “component name structure” element, which permits post-coordination.

Communication of SNOMED CT data using explicit structures for post-coordination is strongly recommended. However, where local agreements permit, other solutions may be used. This is discussed further in the next section.

7.5.3 Mapping to and from message interface semantics

Many communication constructs have a built-in, or assumed semantic model.

Example:

Rather than having a single coded expression to represent a procedure the HL7 Version 3 class "Procedure" contains the following coded attributes.

- Code (cd)
- Priority (priority_cd)
- Reason (reason_cd)
- Method (method_cd)
- Procedure_site (procedure_site_cd)
- Approach_site (approach_site_cd)

Similar constructs occur in other HL7 Version 3 classes (e.g. Observation) and message standards from other sources. However, the HL7 Version 3 "Procedure" example shown here is probably the best example of a particular dilemma for those communicating with a message that takes some aspects of semantics to the structural level while leaving others to the coding scheme.

Suppose we want to communicate the following procedure:

- "Emergency removal of foreign body from stomach by incision"

The HL7 "Procedure" would allow this to be communicated in many ways some of which are shown in Table 7.3.

These first three possible ways of using the structure use the Concepts "Removal of foreign body from stomach by incision" and the priority "Emergency." This is the minimal post-coordinated form, because the complete Concept does not exist in a pre-coordinated form.

- The first option uses the post-coordinated SNOMED CT representation and leaves the other attributes blank.
- The second uses the structure and does not use modifiers.
- The third duplicates the priority both in the structure and in the post-coordinated form.

While the first three options represent the simplest representations, the fourth, fifth and sixth examples adopt the other extreme. There are of course many possibilities in between.

The main point of this illustration is to stress the potential for confusion even when using the same communication structure and the same terminology. This is not a problem of SNOMED CT or entirely of message design. The root of the problem is defining the interface between structural models and semantic models. The message design is in this form largely because of the paucity of terminologies to populate a more rational message.

SNOMED CT represents a massive step forward by providing a consistent semantic model. Further study of the examples in Table 7.3 reveals the strength of SNOMED CT. Using the defining characteristics of SNOMED CT it is possible to assert that these expressions are computably equivalent. The constraint is that where structural fields are used they are used consistently.

To harmonize communication structures and the semantic structures of SNOMED CT, implementation guidelines are needed for specific messages. These guidelines should allow (or explicitly constrain) the use of modifiers that support the full richness of SNOMED CT expressions and should also dictate which of the structural fields should be populated with derived attributes.

HL7 plans to form a working group or special interest group to develop guidelines and examples using HL7 V3 with terminology models, starting with SNOMED CT. It is the intent to harmonize the work of these groups.

Table 7.3: Post-coordination and mapping to the HL7 Procedure class

Ref	Code	Priority	Reason	Method	Procedure_site	Approach_site
1	64550003 + Modifier Attribute = 260870009 Value = 25876001					
2	64550003	25876001				
3	64550003 + Modifier Attribute = 260870009 Value = 25876001	25876001				
4	118828004+ Modifiers x 6 Attribute = 260870009 Value =25876001 A = 260686004 V =129287005 A = 63704007 V = 69695003 A = 260686004 V=129303008 A = 363700003 V = 19227008 A = 260669005 V= 129236007					
5	64550003 + Modifier Attribute = 260870009 Value = 25876001	25876001	19227008	129303008	69695003	129236007
6	118828004+ Modifiers x 6 Attribute = 260870009 Value =25876001 A = 260686004 V =129287005 A = 63704007 V = 69695003 A = 260686004 V=129303008 A = 363700003 V = 19227008 A = 260669005 V= 129236007	25876001	19227008	129303008	69695003	129236007

Concepts used in examples

ConceptID	Meaning	ConceptID	Meaning
118828004	Procedure on gastrointestinal tract	260686004	Method
129236007	Open approach	260870009	Priority
129287005	Incision – action	363700003	Direct morphology
129303008	Removal – action	363704007	Procedure site
19227008	Foreign body	69695003	Stomach structure
25876001	Emergency	64550003	Removal of foreign body from stomach by incision
260669005	Access		

7.5.4 Using Subsets to represent allowable value sets

Standard message specifications and communication agreements with particular user communities often specify restricted lists of codes that can be used in particular message elements:

- Examples of this include the HL7 idea of “vocabulary domains” containing “value sets” specified for use either in a general or specific context in a message element.
- The UK NHS specification for laboratory report messages, which refers to a “bounded list” of Read Codes that are to be used in particular fields of the message.

It is inevitable that a broad terminology such as SNOMED CT needs to be restricted in this way. A message element intended for representation of a “requested radiology investigation” must clearly contain a ConceptID that represents a radiology procedure. The limitations may go further than this. The list of procedures that can be requested may be restricted by local convention or regulation.

Concept Subsets can be used to represent value sets that are permitted in a particular type of message or within a particular user community. This facilitates use of a general-purpose SNOMED CT enabled terminology services to populate and validate the coded elements of messages.

8 Legacy Information Migration Guide

8.1 Introduction

8.1.1 What is migration?

The transition to SNOMED CT from legacy coding systems[15] requires several changes. Many of the most important changes relate to organization and user training, which are outside the scope of this technical guide.

From the technical perspective there are two principal migration issues. The first issue is maintenance of the integrity and value of pre-existing data recorded using other coding schemes (legacy data). The second issue is the maintenance and development of the functionality delivered by software applications that use queries and protocols that include or refer to codes in other coding schemes (legacy queries and protocols).

8.1.2 What will need migrating?

Migration is the process of enabling patient record systems to store, retrieve and communicate clinical information using the new SNOMED CT terminology.

Components that will be affected include:

- Coded data stored in non-SNOMED CT systems
- Information systems, e.g. software and hardware
- Reports
- Decision support
- Protocols

That rely on coded information

8.1.3 Who should be concerned with migration?

The intended audience for this document is any individual or any organization that wishes to develop or use systems that will use SNOMED CT specifically:

- **Clinical software developers**, including those who have worked with a version of the Clinical Terms (the Read Codes) or with SNOMED terminologies.
- **Clinicians**, whose patient data has been stored in non-SNOMED CT systems and who rely on reports and decision support from these systems.
- **Healthcare planners, managers and information specialists** who rely on the secondary use of coded clinical information.

8.1.4 Should legacy codes be overwritten with SNOMED CT Conceptids?

Migration does not mean over-writing legacy data (e.g. SNOP or Read Codes) with SNOMED CT codes. ***This practice is strongly discouraged*** and users are advised to ensure that data stored at the time of data entry is preserved. This is essential for two main reasons:

- Medico-legal status of an altered clinical record may become degraded.
- The original record may be an invaluable resource, should migration produce unexpected results.

[15] Legacy codes refers to any coding scheme used prior to SNOMED Clinical Terms. In the main it means NHS Clinical Terms Version 3, the Read Codes, SNOMED International and SNOMED-RT.

8.1.5 The scale of the task

A substantial body of clinical information resides in electronic systems, represented using existing coding schemes, terminologies and classifications. This information may be of value to individual patient records or to population aggregations. Similarly, there are many queries and decision support protocols that contain knowledge representation based on existing terminologies.

It is not possible to apply a single simple set of rules for migration due to the heterogeneous nature of data storage.

8.1.6 Large user base and databases

In the UK alone, there are about 45 million patients with primary care electronic records. This represents a large amount of coded clinical data. Some systems will be storing many years' worth of patient data, and many hard-coded reports will have been written to operate on this data.

8.1.7 Data quality and consistency

Different users in different settings may select codes and terms in idiosyncratic ways to reflect their needs. While this may be acceptable locally, it can create difficulties if a system developer attempts to provide global solutions for all its customers, as assumptions that apply to one practice or hospital may not apply to others.

8.1.8 Different coding schemes and different versions

Several different coding scheme versions are in use and the process of migration varies for each of these:

- SNOMED
  - SNOMED RT
  - SNOMED International
  - Earlier versions including
    - SNOMED II
    - SNOP
- NHS Clinical Terms Version 3
- The Read Codes
  - Four-Byte
  - Version 2
  - Plus several variants including
    - V1 Unified
    - GP 5 Byte
  - Local coding systems are also in use, often in combination with Read Codes.

8.1.9 Different information systems

There are many system suppliers. As a result of system development and commercial mergers and takeovers, many suppliers support more than one application in the same domain. The challenge is to migrate from this diverse situation to a next generation environment supporting standards such as SNOMED CT.

8.1.10 Different record architectures

There are also many different record structures and architectures within these systems.

8.2 General Advice on Management of Legacy Data

8.2.1 Strategies for data migration

Moving from a legacy coding scheme to SNOMED CT requires attention to be paid to continued accessibility and use of data encoded using the legacy scheme. Three options should be considered:

- Mapping or converting the data:
  - This requires all the legacy-coded data to be mapped to SNOMED CT.
  - If this is done the legacy codes should also be retained to ensure that information can be recovered in the event of inappropriate mapping or if improved maps become available in future releases of SNOMED CT.
- Linking or integrating the data in ways that allow legacy coded data and SNOMED CT encoded data to be retrieved by a common set of queries generated using same tools and queries. This approach supports the co-existence of unconverted legacy data with current data. There are two requirements for this:
  - The legacy codes or identifiers must be recognizably different from the data encoded using SNOMED CT.
    - This may be achieved using a coding scheme identifier. However, the design of SNOMED CT identifiers ensures that SNOMED International codes, Read Codes and ConceptIDs are all distinguishable.
  - The relationship between Concepts in SNOMED CT and those in the legacy coding scheme must be recognized when retrieving data.
    - This is possible for combinations of SNOMED CT, SNOMED International, Clinical Terms Version 3, Read Codes Version 2 and Read Codes 4-Byte Set for the following reasons:
      - The Concepts Table contains the appropriate legacy codes
      - The Relationships Table contains historical “same as,” “may be a” and “replaced by” Relationships which link Inactive Concepts to the current SNOMED CT subtype hierarchy.
- Archive or retain old data in its original form, and where it is necessary to retrieve historical information, use components from the legacy system to do this.
  - This approach completely separates the new SNOMED CT from the legacy data and is unlikely to be acceptable in clinical practice. However, it may be appropriate for some data warehousing applications where the wholesale conversion of data is considered too onerous.

8.2.2 Overview of methods for migrating legacy data

The previous section discusses three possible strategies for dealing with legacy data. The following notes summarize a general method for pursuing one of these options, which involves mapping of legacy data to SNOMED CT representations.

The overall intention of this method is that:

- Legacy codes stored in records are mapped to ConceptIDs.
- Legacy terms or term identifiers stored in records are mapped to DescriptionIDs (optional).
- The original encoded data is retained and remains accessible to reuse for future updating if improved mappings become available in future SNOMED CT releases.

The developer must have a good understanding of the structure and semantics of the legacy data. This includes being aware of:

- The coding scheme version(s) used to encode the data.
- Whether term strings and/or term identifiers are stored.
- Any semantics implied by record structures or proprietary modifier, which may be inadvertently lost in mapping.

The first step in the actual mapping is to find the Concept that contains each legacy code in the stored records. This allows the directly matching Concept to be found. The search for this Concept depends on the legacy coding scheme, as indicated below:

- SNOMED International
  - Search for a precise match in the SNOMEDID field
- Read Codes Version 2 and Clinical Terms Version 3
  - Search for a precise match in the CTV3ID field.
- Read Codes Four-Byte Set
  - Search for the CTV3ID field for the four-byte code prefixed by a full-stop.
  - Thus, if the four-byte code is “H43.” Search for the five byte code “.H43.” in the CTV3ID field.

Check the ConceptStatus of the matched Concept and either retain the ConceptID or replace it as follows:

- ConceptStatus = “Duplicate”
  - The ConceptID may be replaced by the ConceptID referred by the “same as” Relationship of the matched Concept.
- ConceptStatus = “Ambiguous”
  - The ConceptID may be replaced by the ConceptID referred by one of the “may be a” Relationships of the matched Concept. However, this should only be done if a stored term (or term identifier) is available which allows manual or automatic disambiguation. The Clinical Terms Version 3 DCF file is a resource that may assist in automatic disambiguation.
- ConceptStatus = “Erroneous”
  - The ConceptID may be replaced by the ConceptID referred by the “replaced by” Relationship of the matched Concept.

If the ConceptStatus is not one of these values or if no appropriate Relationship is found retain the ConceptID.

- Note that it is not essential to replace the ConceptIDs of Inactive Concepts in the manner specified above. As an alternative Inactive Concepts can be included in selective retrieval using methods discussed in section [7.4.4]. However, replacement may enhance performance and, so long as the legacy code is also retained, the matched inactive ConceptID provides no added value.

If the application also stores DescriptionIDs and the legacy data contains the text or identifiers of terms, consider on of the following approaches for deriving DescriptionIDs from the legacy data:

- Check the term stored with the legacy data against the Descriptions associated with the mapped Concept.

  or

- Use the legacy coding system to lookup the term associated with the stored term identifier and check this against the Descriptions associated with the mapped Concept.
    - In the case of Read Codes Four-Byte Set and Read Codes Version 2 this process may be enhanced by using the term mapping file.

After mapping quality review is recommended. Possible methods of review include:

- Checking that the supertype ancestors[16] of Concepts in mapped records are consistent with their use in the stored record (or the role of the original code in the legacy system).
    - For example, this approach can be used to confirm that a procedure record does not contain a mapped ConceptID for a disorder or finding.
- Running reports or queries, which have known results when applied to the legacy data. The results with mapped data may not be identical due to improvements in the specificity and completeness of reporting. However, any discrepancies provide a useful focus for manual review.
- Manual review of samples of the mapped data should also be considered.

8.3 Specific Assistance for Users of Legacy SNOMED Systems

8.3.1 Migration from SNOMED RT

Migration from SNOMED RT poses very few significant issues, since many features of the design of SNOMED CT including the use of SCTIDs were incorporated into SNOMED RT.

The transition to SNOMED CT for users of SNOMED RT is relatively straightforward because the ConceptIDs of SNOMED RT are for the most part the same as those used in SNOMED CT. In some cases, during the merger of SNOMED RT and Clinical Terms Version 3 some Concepts in SNOMED RT have been found to be ambiguous or duplicated. These Concepts have been inactivated by an appropriate change of ConceptStatus but are still present in the Concepts Table and are linked to Active Concepts by the following historical relationships:

- Each duplicate Concept has a “SAME AS” Relationship to the Active Concept with the same meaning as the duplicate Concept.
- Each ambiguous Concept has “MAY BE A” Relationship to one or more Active Concepts, which represent possible disambiguated meanings.

These Relationships can be used either to allow Concepts recorded using these Concepts to be recognized by retrieval tools or to enable mapping of the stored information to the appropriate active ConceptID.

If any stored ConceptID of an Inactive Concept is mapped to an active ConceptID using these Relationships it is strongly recommended that the original ConceptID is also retained. This enables future improvements or corrections of such mappings if revised Relationships are present in a future release of SNOMED CT.

In addition, SNOMED RT contained both generic and brand name drugs for the US. A decision was made during the merger process to not retire these concepts using the extension mechanisms, but to place these components directly in the US Drug Extension. Therefore to access all SNOMED RT Components you will need to use the US Drug Extension in addition to the SNOMED CT International Release.

[16] Subtype tests are discussed in [6.2.5].

Like SNOMED CT, SNOMED RT also contains the appropriate legacy SNOMEDID from SNOMED International.

8.3.2 Migration from SNOMED International

The meaning of coded clinical data encoded using SNOMED International is maintained using mechanisms that support concept permanence and version control.

Concept permanence ensures that codes assigned in SNOMED International are retained, accessible, and not reused. The codes used in SNOMED International are present in SNOMED CT in the SNOMEDID field of the Concepts Table. Even when a Concept is retired from active use its code is never reassigned to another Concept.

The SNOMEDID in the Concepts Table can be used either to allow recognition of legacy data by SNOMED CT retrieval tools or to enable mapping of the codes and storage of the appropriate ConceptID.

8.3.3 Migration from earlier versions of SNOMED

To assist in the migration of legacy data from early version of SNOMED – SNOMED II (1979) and SNOP (1965), a Bridge File (or mapping table) is available from the IHTSDO which links the legacy code to its corresponding code in SNOMED International.

As noted in the previous section, SNOMED CT retains the codes used in SNOMED International in the SNOMEDID field of the Concepts Table. This can be used to complete the linkage or mapping of legacy data encoded using earlier versions of SNOMED.

8.4 Specific Assistance for Users of Read Coded Systems

8.4.1 General issues

Migration from Clinical Terms Version 3 is a little more difficult than migration from SNOMED RT as there are changes to the format of some identifiers and some structural changes in representation. However, most of the features of SNOMED CT are similar to those of Clinical Terms Version 3.

Migration to SNOMED CT from early versions of the Read Codes presents five significant changes:

- Size and complexity
- Meaningless concept identifiers
- Multiple hierarchies
- Post-coordination
- Hierarchy (IS_A relationships) changes.

These changes confer considerable advantages but the impact of these changes should be taken into account when designing SNOMED CT systems or reusing data stored using legacy codes. These issues are discussed in the following section.

8.4.2 Issues specific to migration from early Read Code versions

Size and complexity

SNOMED CT is a much larger terminology and has much greater coverage and specificity than the early versions of the Read Codes. There are over 350,000 Concepts and a million Relationships.

It also has a more complex structure to allow support for the

- Semantic relationships between Concepts.
- Language Subsets to support dialects and multi-lingual implementations.
- Navigation Subsets to allow alternative hierarchical view of the terminology.
- Other Subsets to support particular countries and specialties.

Meaningless concept identifiers

The use of 'meaningless' identifiers in SNOMED CT indicates that ConceptIDs cannot be used to determine the nature of a concept.

Examples:

- In Read Codes Version 2 all 'Surgical operations on the ear' can be retrieved by selecting all Read Codes starting with '73'
- In SNOMED International all 'Diagnoses related to the digestive system' can be retrieved by selecting all codes starting with 'D5'

In SNOMED CT the hierarchical information about the Concept is in the Relationships Table. This change allows support for multiple hierarchies.

Multiple hierarchies

The early versions of the Read Codes place each Concept in only one hierarchy, based on the sequence of characters in the code. However, many clinical concepts are logically subtypes of several other concepts. SNOMED CT (like Clinical Terms Version 3) supports this through a set of explicitly stated relationships.

Example:

The term 'Kidney biopsy' has a parent of 'Biopsy of urinary system' AND a parent of 'Procedure on kidney.'

Post-coordination

Overview

SNOMED CT, like Clinical Terms Version 3, allows post-coordination of multiple ConceptIDs to specify a more complex or detailed Concept. Several examples of this are discussed in the following sections.

Context

Example:

'*Family history* of asthma'

This concept is encapsulated in a single code in earlier versions of Read Codes but can be post-coordinated in SNOMED CT.

Terminology	Code	Term
Four-Byte Read Codes	12D2	'Family history of asthma'
Read Codes Version 2	12D2.	'Family history of asthma'
Clinical Terms Version 3	H33..	Asthma This assumes some other mechanism for representation of context.
SNOMED CT	281666001 + (246090004=195967001)	'Family History of disorder' qualified by 'Associated finding' = 'Asthma'
	315823007	'Family history of asthma' is also available as a pre-coordinated value.[17]

Qualifiers

Example:

'*Emergency* appendectomy'

This concept is encapsulated in a single code in Read Codes Version 2, whereas in Clinical Terms Version 3 and SNOMED Clinical Terms this can be represented in a post-coordinated fashion.

a. Terminology	b. Code(s)	c. Term(s)	d. Defined
Read Codes Version 2	7700.	'Emergency appendicectomy'	Pre-coordinated
Clinical Terms Version 3	7700.	'Emergency appendicectomy'	Pre-coordinated
	X20Wz + (X904E= X78uH)	'Appendicectomy' qualified by 'Priority' = 'Emergency'	Post-coordinated
SNOMED CT	174036004	'Emergency appendectomy'	Pre-coordinated
	80146002 + (260870009=25876001)	'Appendectomy' qualified by 'Priority' = 'Emergency'	Post-coordinated

[17] Note that in SNOMED CT where pre- and post-coordinated alternatives exist, the pre-coordinated expressions will have defining characteristics that match the post-coordinated form of expression. In the early releases of SNOMED CT this will only be true for the released Relationships Types. For this reason, some situation with explicit context Concepts (such as family history) may initially not have computable equivalence with their post-coordinated representations.

Hierarchy 'changes'

Reports based on legacy concepts and hierarchies may not produce comparable results when migrated to SNOMED CT, as this example from the Read Codes shows.

Example:

If a query has been designed to retrieve all recorded cases of asthma, based on the parent Read Code in Version 2 (H33..), then the terms returned would be:

Read Codes Version 2

H3…	Chronic obstructive lung disease	
	H30..	Bronchitis unspecified
	H31..	Chronic bronchitis
	H32..	Emphysema
	H33..	Asthma

However, if the same query is used on a Clinical Terms Version 3 – based system, a different set of concepts will be included, because the hierarchy has changed. Asthma is no longer an immediate descendant of 'H3…', but instead a child of 'H…. *Respiratory Disorder*.'

Clinical Terms Version 3

H….	Respiratory disorder		
	H33..	Asthma	
		X101x	Allergic asthma
		XE0YT	Non-allergic asthma
		X1023	Drug-induced asthma
		173A.	Exercise-induced asthma

In SNOMED CT the meaning of the Concept "asthma" is the same as for Clinical Terms Version 3, but its descendants may differ.

SNOMED Clinical Terms

50043002	Disease of respiratory system (disorder)		
	1063349207	Asthma	
		225057002	Brittle asthma
		233679003	Late onset asthma
		266361008	Intrinsic asthma
		266364000	Asthma attack

A query developed for Clinical Terms Version 3 may have to be adapted to generate comparable output for SNOMED CT. However, since SNOMED CT represents a more complete subtype hierarchy a SNOMED CT query will in most cases be both more accurate and comprehensive. This means it will include a more complete set of subtypes of the Concept asthma and will be less likely to include Concepts that are not subtypes of asthma.

8.4.3 Clinical Terms Version 3 as a Superset of the Read Codes

In 1997, Clinical Terms Version 3 (The Read Codes) was made a superset of all previous Read Code versions. This means that all the codes used in the Read Codes Four-Byte Set and Read Codes Version 2 can also be found in Clinical Terms Version 3.

All subsequent releases have been maintained as part of the Clinical Terms Version 3 Superset. This feature will assist migration to SNOMED CT by allowing a harmonized approach to be applied to all versions.

Some of early version codes regarded as “redundant” or “extinct” in Clinical Terms Version 3 are accessible through a resource called the “Description Change File” (DCF). The SNOMED CT equivalents of these states are the ConceptStatus values “duplicate” and “ambiguous”. The relevant Concept mapping information has been carried forward from Clinical Terms Version 3 using the “SAME AS” and “MAY BE A” Relationships. Additional resources are available in Clinical Terms Version 3 for mapping terms between versions. These mappings are considered in the Technical Reference Guide.

8.4.4 Read Code Term mapping resources

Introduction

This section describes additional resources available from the UK NHS for enhancing and the mapping of terms and concepts encoded using Read Codes Four-Byte Set and Read Codes Version 2.

Term Mapping Files

A Clinical Terms Version 3 term identifier (TermIdD) can be generated for every Read Code and term in the Four-Byte Set, and for every Read Code, TermCode, and term in Version 2. The TermID is a five-character term identifier as shown below, and the term’s text string can be derived from the Clinical Terms Version 3 Terms file (terms.v3).

Two term mapping files are used in England:

- Four-Byte Term Mapping File (Fbtermv3.v3)
- Version 2 Term Mapping File (V2termv3.v3)

Two alternative mapping files are used in Scotland:

- Four-Byte Term Mapping File (Fbtmscot.v3)
- Version 2 Term Mapping File (V2tmscot.v3)

Four Byte Term Mapping File

Four-Byte Read Code	Four-Byte Term	V3 TermID
F111	Bacterial meningitis	Y009Y

Version 2 Term Mapping File

Version 2 Read Code	Term Code	V2 Term	V3 TermID
G3…	00	Ischaemic heart disease	Y201T

The Description Change File (DCF)

The Description Change File (DCF) provides mappings of all Read Codes and terms from all versions of Read Codes to the most up-to-date assignment of a Read Code to each term in Clinical Terms Version 3.

Four Byte Read Codes

All four Byte Read Codes are redundant in Clinical Terms Version 3, with a persisting Read Code provided. The Four-Byte Read Code has a preceding dot added in Clinical Terms Version 3.

Example:

	Four-Byte Read Code	V3 Read Code	
		Redundant	Persisting
Bronchoscopy	3624	.3624	X00qf
Hysterectomy	7CB.	.7CB.	X403B

The DCF represents this as follows:

Old Code	New Code	TermId	Map Status	Date
.3624.	X00qf	Y021D	R	1997-10-01
.7CB.	X403B	Y40D5	R	1997-10-01

Therefore, the concept in the Four-Byte Set with Read Code '3624' is represented as a redundant Read Code '.3624', which has a persisting Clinical Terms Version 3 Read Code 'X00qf'. The term string in the Four-Byte Set is represented by the Clinical Terms Version 3 TermID 'Y021D'. The MapStatus indicates that this mapping is due to the source concept being redundant.

Version 2 Read Codes

Version 2 Read Codes are integrated into Clinical Terms Version 3 in two ways. Most codes are integrated directly. Note that these codes have a trailing dot.

Example:

	V 2 Read Code	V 3 Read Code
Leukoplakia of cervix	K552.	K552.
Hypertensive heart disease	G21..	G21..

Therefore, the Version 2 codes are re-used in Clinical Terms Version 3. The remaining codes are integrated indirectly via the Description Change file in a similar fashion to the Four-Byte set where duplication of concepts occurred in Version 2.

Example:

Tuberculous meningitis has two Read Codes in Version 2. In Clinical Terms Version 3 it has only one current/optional code (F004).

	V 2Read Code	V 3 Read Code
Meningitis – tuberculous	F004.	F004.
Tuberculous meningitis	A130.	F004.

The DCF represents this as follows:

Old Code	New Code	TermID	Maps Status	Date
A130.	F004.	Y009z	R	1997-10-01

(‘Y009z’ is the TermID for Tuberculous meningitis)

Ambiguous concepts: a special case

As previously stated, earlier versions of Read Codes have structural limitations, and one consequence of this feature is that some of the concepts in these versions are inherently ambiguous.

Example:

Read Code Version 2 ‘G30..’

Read Code	Term Code	Term	Status
G30..	00	Acute myocardial infarction	Pref
G30..	11	Attack – heart	Syn
G30..	12	Cardiac rupture after acute myocardial infarction	Syn
G30..	13	Heart Attack	Syn
G30..	14	MI – Acute myocardial infarction	Syn
G30..	15	Silent myocardial infarction	Syn
G30..	16	Coronary thrombosis	Syn

As can be seen, this concept in Version 2 a number of impure synonyms e.g. ‘Cardiac rupture after acute myocardial infarction’. Clinical Terms Version 3 indicates the ambiguity of this concept by marking the code “G30..” as “extinct.” However, the constituent elements of this ambiguous concept (e.g. “Silent myocardial infarction”, “Heart attack”) are perfectly valid and so the DCF assists in the migration by pointing Codes to the most up-to-date assignment of a Read Code to each term in Clinical Terms Version 3.

Example:

‘Acute myocardial infarction’

The DCF represents this as follows:

Old Code	New Code	TermID	Maps Status	Date
G30..	XEOUh	Y202N	R	1997-10-01

Therefore, the precision of correct assignment of a Clinical Terms Version 3 code is enhanced if the term string has been preserved in the legacy system.

What if term strings have not been stored in the legacy system?

Some systems may not have stored term strings in their records but simply stored the Read Codes alone. While this may not be a problem, in many cases it does limit the precision of the mapping process.

9 Extension Services Guide

9.1 Rationale for Extensions

An Extension mechanism allows authorized organizations to add locally valid content and Subsets without compromising the main body of SNOMED CT. This facility will be valuable to:

- Meet the needs of specialties and realms.
- Meet vendor needs.
- Meet local business needs.

9.2 Installing Extensions

Extensions follow the file structures used for the main body of SNOMED CT. Therefore, the process of importing an Extension is similar to the core data import process (see section [6.4]).

The main technical difference is that Extensions are added to, rather than replace, data in the SNOMED CT tables (see [6.4.3]).

Before installing any Extension it is important to check that:

- The Extension has been supplied by the IHTSDO or another organization authorized by the IHTSDO to provide such Extensions.
- You are satisfied with the quality control procedures of the providing organization.
  - Authorization of an organization to produce Extensions does not imply any seal of approval related to the quality of Extensions provided by those organizations.
  - Installation of Extensions is done entirely at the risk of the user subject to their license agreement with the provider of the Extension and/or the application developer.
- Any dependencies of the Extension have been met. These dependencies may include
  - Installation of a particular SNOMED CT release;
  - Prior installation of other Extensions.
- The installation procedure has pre-checked all Components in the Extension to ensure that:
  - All ComponentIDs
    - Are unique.
    - Have partition-identifier and namespace-identifier values appropriate to the provider of that Extension.
    - Have a valid check-digit.
  - All fields meet data type, size and value constraints specified for the relevant tables.

Caution!:

If any Components fail any of these tests the entire Extension must be rejected.

9.3 Editing and Maintaining Extensions

A future document will provide advice for organizations that are authorized to develop Extensions. To obtain a namespace identifier, contact the IHTSDO.

Appendix A: Distribution Table Technical Summary

A.1 Introduction

The tables in this section provide technical details that could be used to develop a database schema for representing all SNOMED CT distribution files. The tables in this section provide a limited set of technical information about each of the fields in the SNOMED CT distribution files. The tables also make recommendations about additional keys, which may be necessary to support an effective implementation in a relational database without restructuring the distributed data. However, this information is advisory only and is not a specification of requirements or a recommendation. Alternative approaches based on restructuring the distributed data may provide a more efficient and responsive method of implementation.

Refer to the specification documents for details about the use and possible values of each field.

A.2 Information in the Tables

A.2.1 Type

One of the five possible data types used in SNOMED CT distribution files:

- SCTID (SNOMED Clinical Terms Identifier)
- String – A text string
  - Represented as a sequence of UTF-8 encoded characters
- Integer – An integer
  - Represented as a sequence of digits encoded as UTF-characters
- Enum – An integer which refers to a list of enumerated options
  - Represented as a sequence of digits encoded as UTF-characters
- Boolean – A Boolean value (true or false)
  - Represented as “0” (false) or “1” (true) encoded as a UTF-8 character

A.2.2 Size

The size of the data represented as:

- The number of bits in an integer representation
  - This is not applicable to the string data type.
  - Possible values 1, 8, 16, 32 or 64.
- The maximum number of characters required for a string representation
  - This is the maximum character length in the distribution file.

A.2.3 Partition

This is only applicable to the SCTID data type. It specifies the partition (or in some cases set or range of possible partitions) of the SCTIDs used in that field.

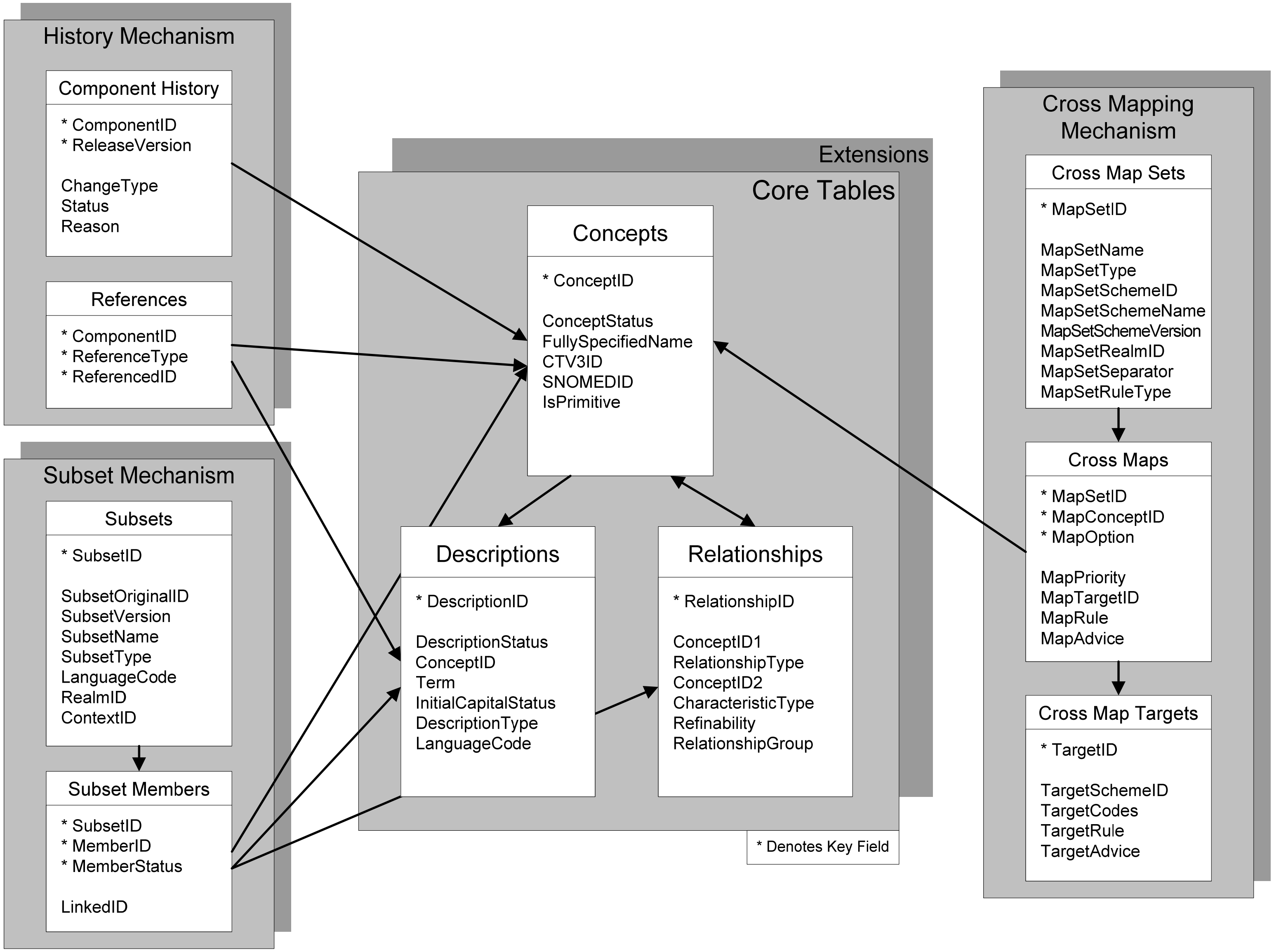


Figure A.1: SNOMED Clinical Terms Data Structure

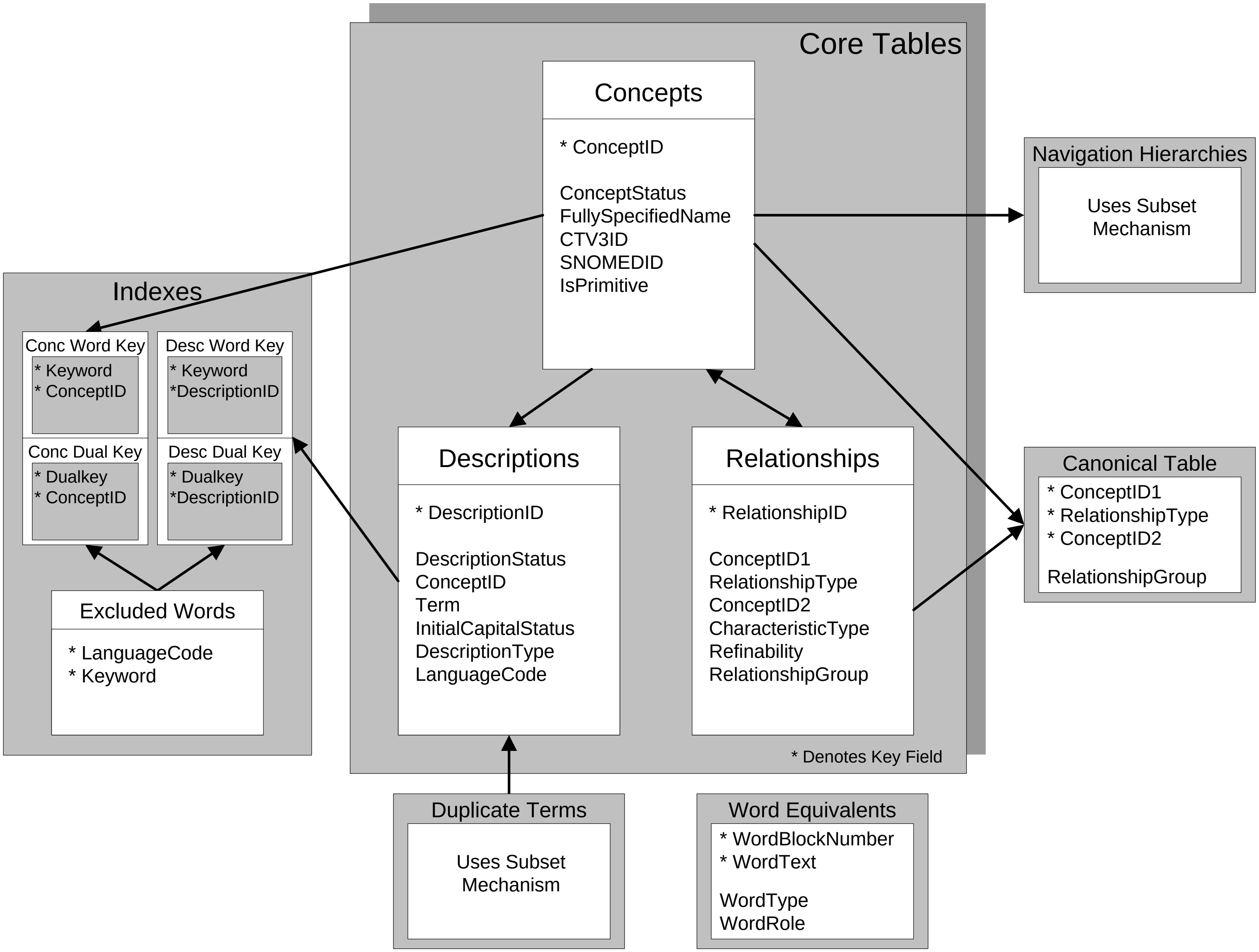
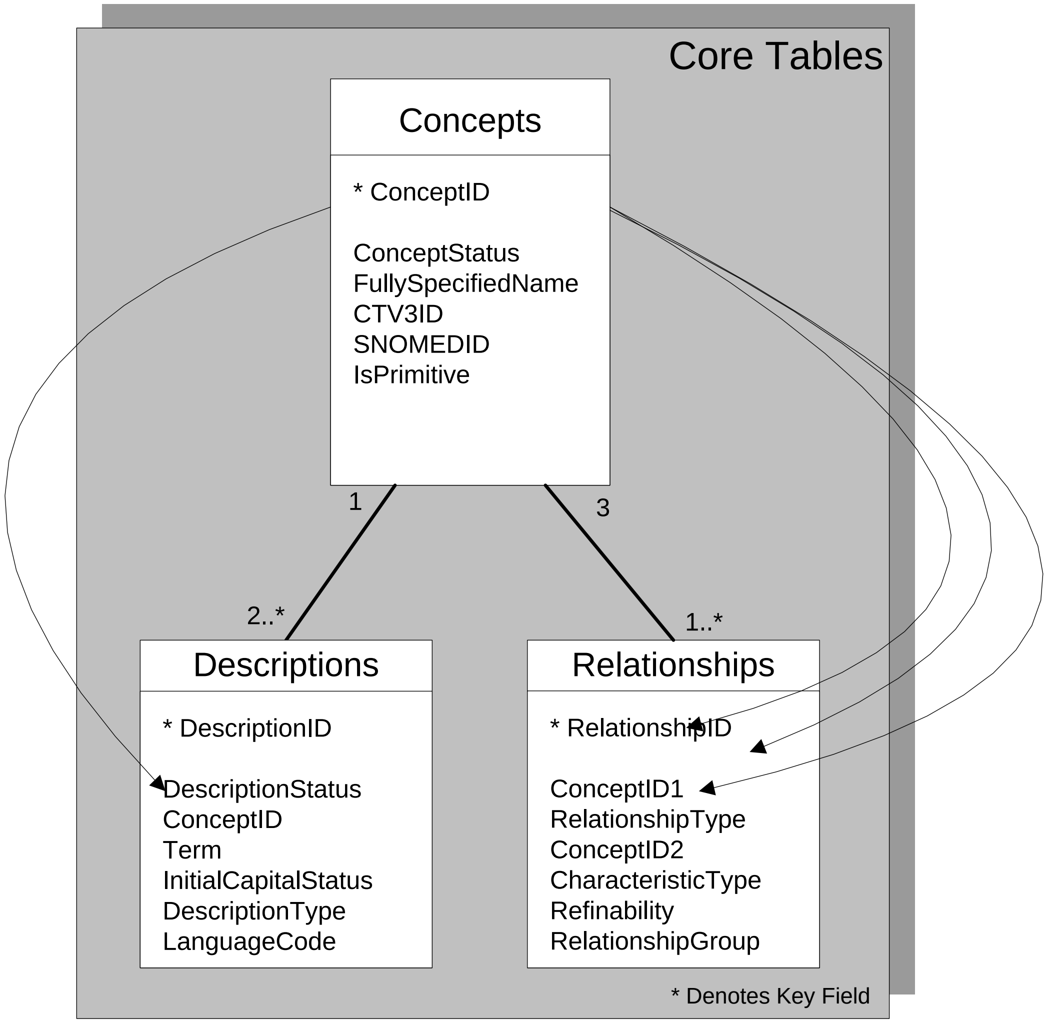


Figure A.2: SNOMED Clinical Terms Data Structure



A Concept is described by the Term of 2-n Descriptions.

Each Description refers to 1 Concept.

A Relationship refers to 3 Concepts: a source, target, and relationship type.

A Concept is the source of 1-n Relationships (except the root Concept).

A Concept is the target of 0-n Relationships.

A Concept represents the type of Relationship.

Figure A.3: The SNOMED CT Core Tables

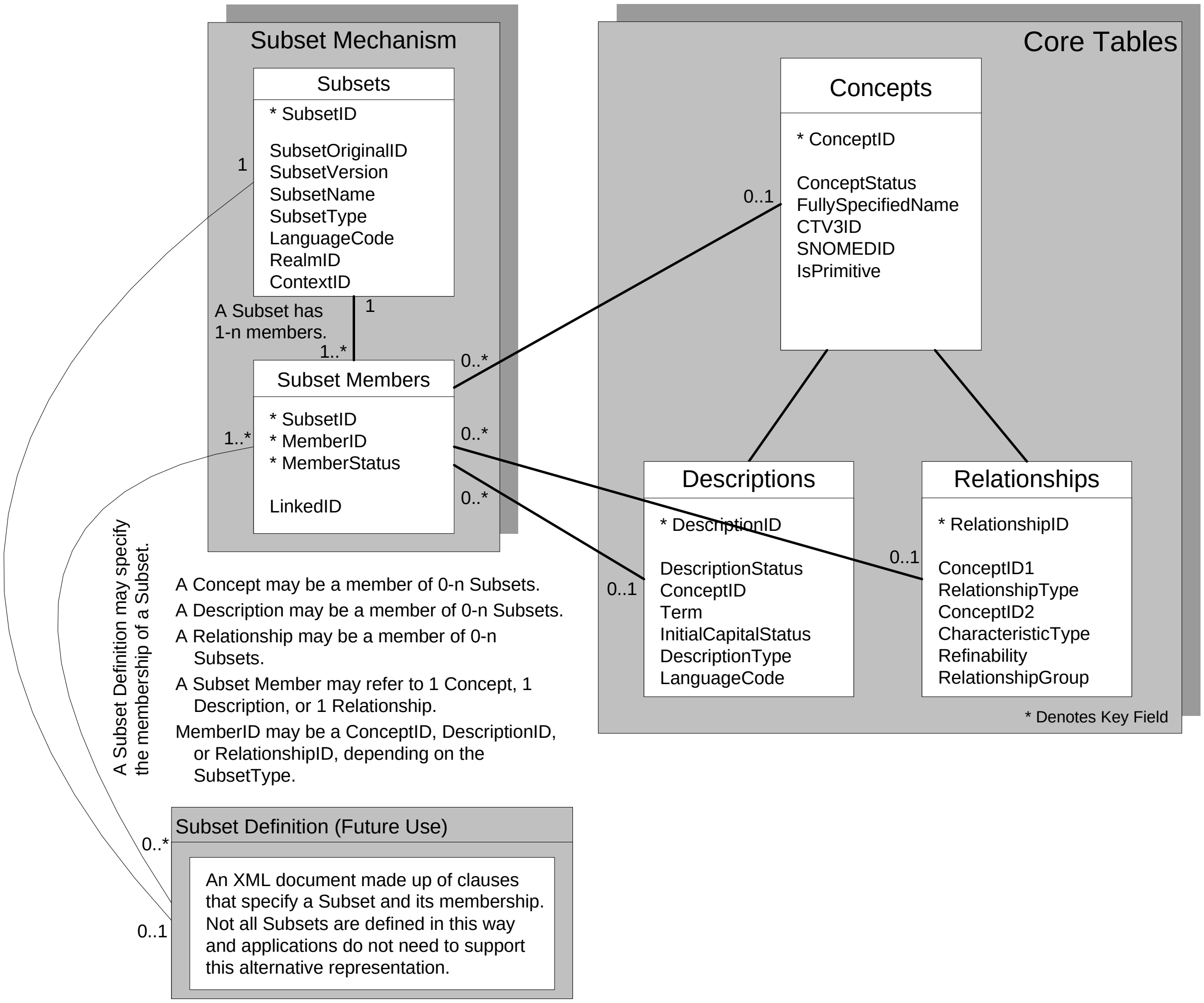


Figure A.4: The SNOMED CT Subset Mechanism Structure

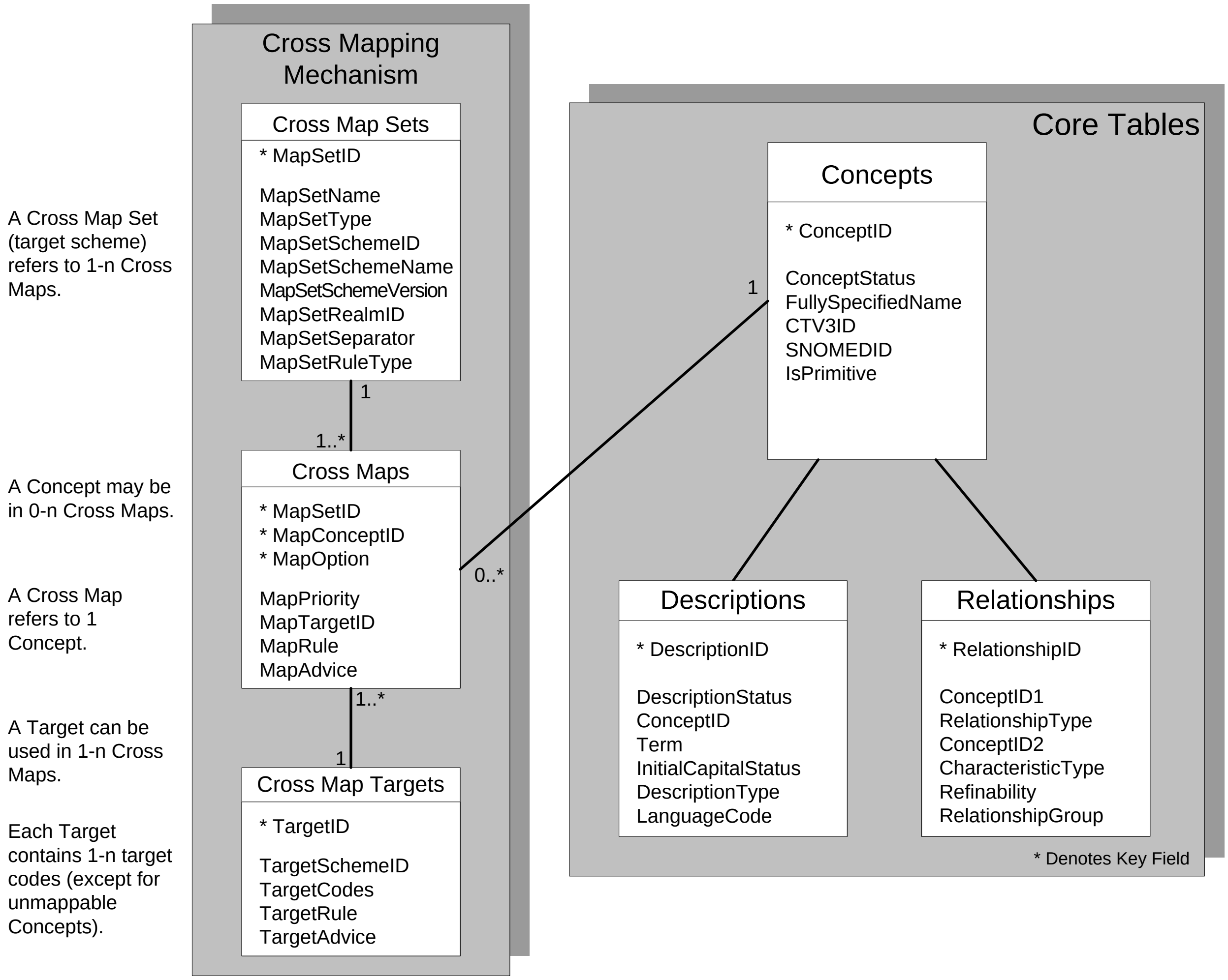


Figure A.5: The SNOMED CT Cross Mapping Mechanism Structure

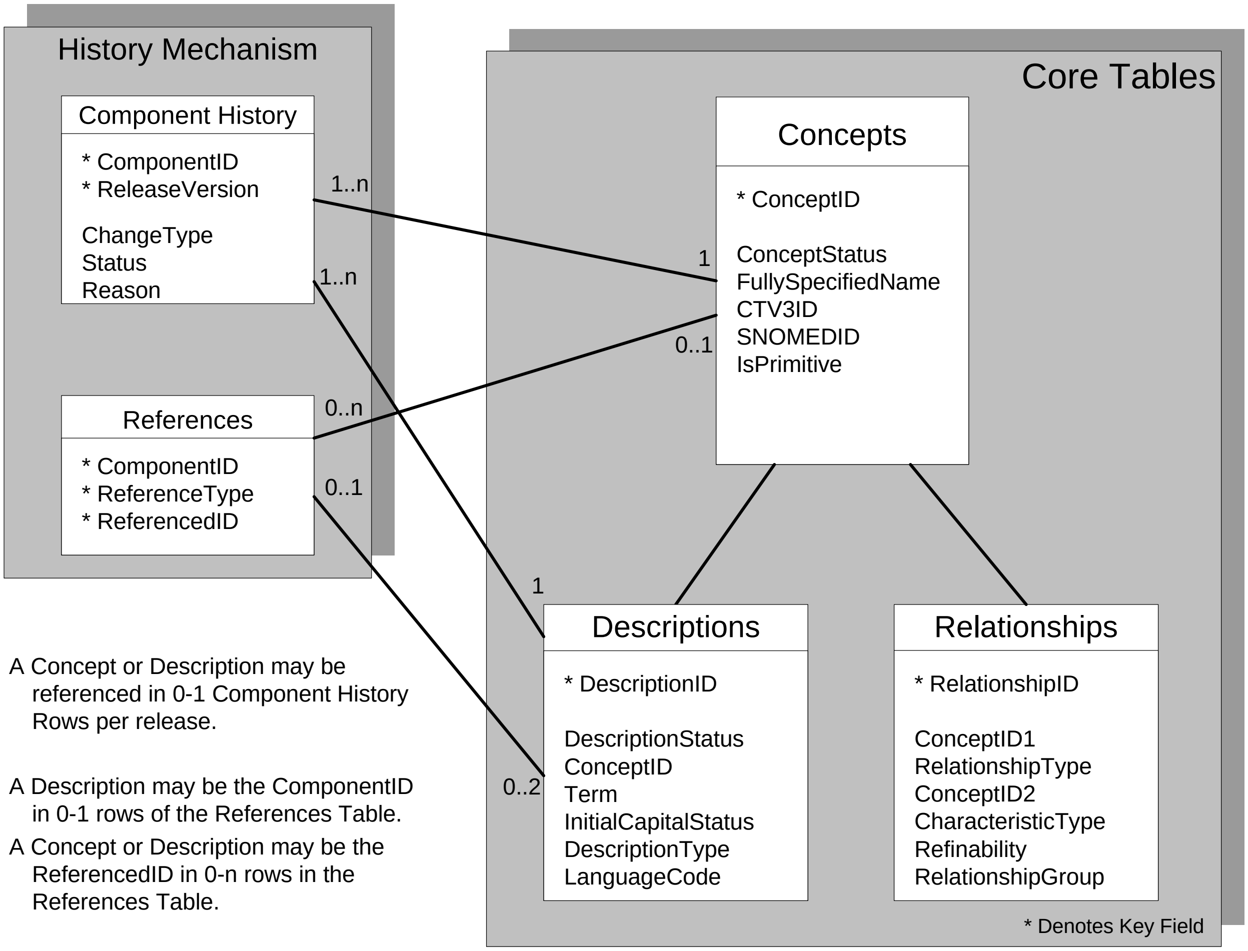


Figure A.6: The SNOMED CT History Mechanism Structure

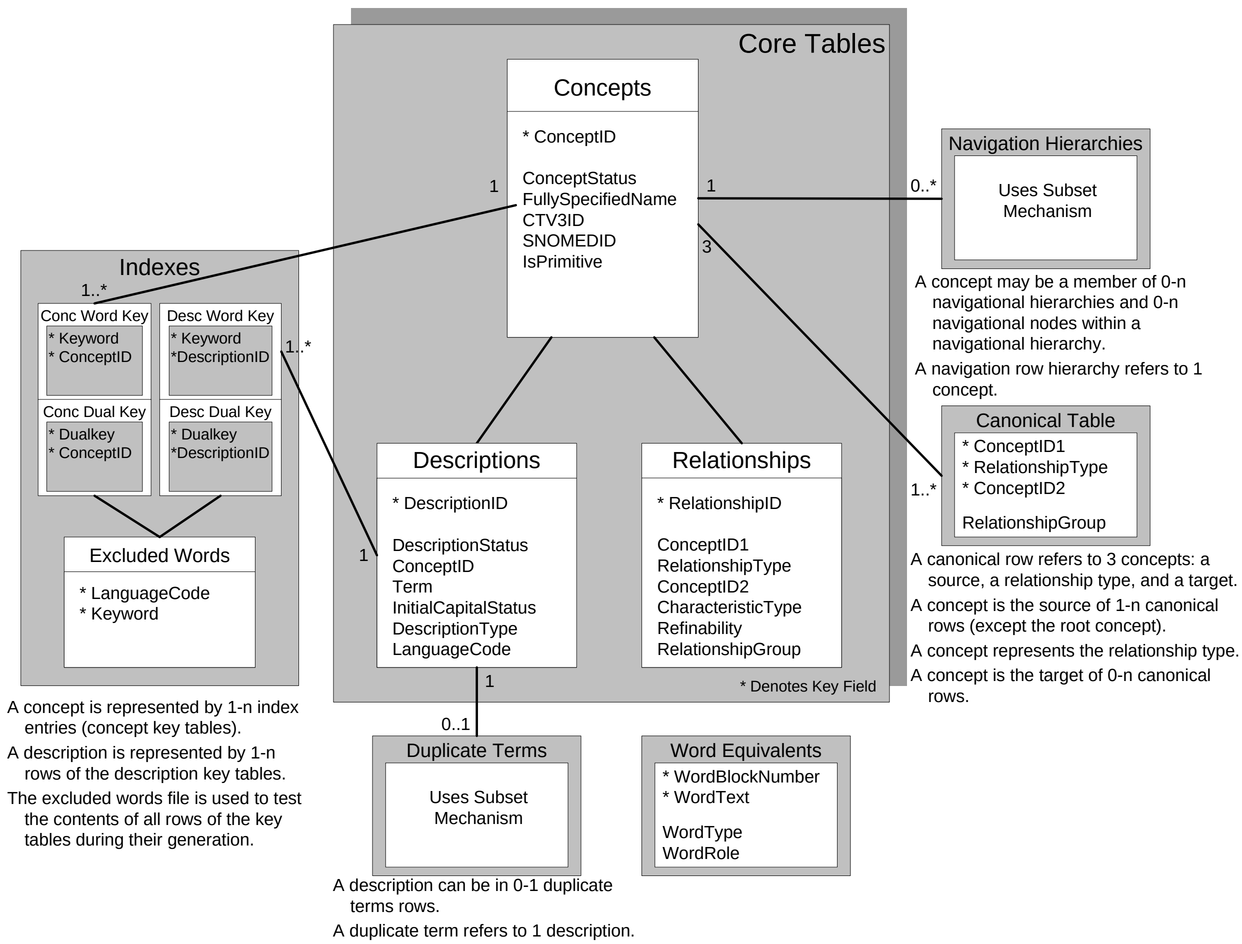


Figure A.7: The SNOMED CT Developer Toolkit Structure

A.2.4 Keys

Colored blocks in the keys column identify keys that may be required or useful for effective implementation:

- Primary keys (blue)
- Recommended keys:
  - Critical fields (red) are required for the relevant functionality.
  - User fields (green) are not essential but may improve functionality.
- Foreign keys used in joins to other tables (purple)

A number is used to indicate the order of fields in combined keys. A letter in the primary key identifies this as a target for a foreign key pointer. The same letter in a foreign key indicates a join based on the primary key identified by the same letter. Foreign keys marked “X” may refer to any component with an SCTID primary key.

Table A.1: Core Tables

Table and field names	Type	Size		Partition	Indices				Use of recommended additional indices
		Integer bits	String length		Primary	Recommended		Foreign	
Concepts	Table	-	-	-		A	B		
ConceptID	SCTID	64	18	0	C				A: To access Concept by CTV3ID (Read Code) B: To access Concept by old style SNOMEDID.
ConceptStatus	Enum	8	2	-					
FullySpecifiedName	String	-	255	-					
CTV3ID	String	-	5	-		1			
SNOMEDID	String	-	8	-			1		
Is Primitive	Boolean	1	1	-					
Descriptions	Table	-	-	-		C			
DescriptionID	SCTID	64	18	1	D				C: To access Descriptions for a specified Concept.
DescriptionStatus	Enum	8	2	-					
ConceptID	SCTID	64	18	0		1		C	
Term	String	-	255	-					
InitialCapitalStatus	Boolean	1	1	-					
DescriptionType	Enum	8	1	-					
LanguageCode	String	-	8	-					
Relationships	Table	-	-	-		D	E		
RelationshipID	SCTID	64	18	2	R				D: To access all or selected Relationships of a specified source Concept. E: To access all or selected Relationships of a specified target Concept.
ConceptID1	SCTID	64	18	0		1		C	
Relationship Type	SCTID	64	18	0		4	3	C	
ConceptID2	SCTID	64	18	0		5	1	C	
Characteristic Type	Enum	8	1	-		2	2		
Refinability	Enum	8	1	-					
RelationshipGroup	Integer	16	5	-		3			

Table A.2: Subset Tables

Table and field names	Type	Size: Integer bits	Size: String length	Partition	Indices: Primary	Indices: Recommended	Indices: Recommended	Indices: Foreign	Use of recommended additional indices
Subset	Table	-	-	-		F			
SubsetID	SCTID	64	18	3	C				F: To access previous versions of the same Subset.
SubsetOriginalID	SCTID	64	18	3		1		S	
SubsetVersion	Integer	16	5	-		2			
SubsetName	String	-	255	-					
SubsetType	Enum	8	2	-					
LanguageCode	String	-	8	-					
RealmID	String	-	24	-					
ContextID	String	-	18	-					
Subset Members	Table	-	-	-		G	H		
SubsetID	SCTID	64	18	3	1	1	1	S	G: To access all members of Subset with the same MemberStatus. H: To access inverse (child to parent) Navigation Links in a Navigation Subset.
MemberID	SCTID	64	18	0, 1, 2	2			X	
MemberStatus	Integer	16	5	-	3*	2			
LinkedID	SCTID	64	18	0			2	C	

Table A.3: History Tables

Table and field names	Type	Size: Integer bits	Size: String length	Partition	Indices: Primary	Indices: Recommended	Indices: Recommended	Indices: Foreign	Use of recommended additional indices
Component History	Table	-	-	-					
ComponentID	SCTID	64	18	0-5	1			X	
ReleaseVersion	Integer	32	8	-	2				
ChangeType	Enum	8	1	-					
Status	Enum	8	2	-					
Reason	String	-	255	-					
References	Table	-	-	-		G	H		
ComponentID	SCTID	64	18	0-5	1			X	G: To access all Components that referenced a specified active Component.
ReferenceType	Enum	8	2	-	2				
ReferencedID	SCTID	64	18	0-5	3	1		X	

* The MemberStatus is only required to be part of the primary key to the Subset Members Table in the case of Navigation Subsets. In all other cases, the combination of SubsetID and MemberID must be unique.

Table A.4: Cross Mapping Tables

Table and field names	Type	Size		Partition	Indices				Use of recommended additional indices
		Integer bits	String length		Primary	Recommended		Foreign	
Cross Map Sets	Table	-	-	-					
MapSetID	SCTID	64	18	4	M				**J:** To access the Cross Map Set for a particular alternative coding scheme.
MapSetName	String	-	255	-					
MapSetType	Enum	8	2	-					
MapSetSchemeID	String	-	64	-		1		OID	
MapSetSchemeName	String	-	255	-					
MapSetSchemeVersion	String	-	12	-					
MapSetRealmID	String	-	24	-					
MapSetSeparator	String	-	1	-					
MapSetRuleType	Enum	8	2	-					
Cross Map Targets	Table	-	-	-					
TargetID	SCTID	64	18	5	T				
TargetSchemeID	String	-	64	-					
TargetCodes	String	-	255	-					
TargetRule	String	-	255	-					
TargetAdvice	String	-	255	-					
Cross Map	Table	-	-	-					
MapSetID	SCTID	64	18	4	1	1		M	**K:** To access maps in a stated priority order.
MapConceptID	SCTID	64	18	0	2	2		C	
MapOption	Integer	16	5	-	3				
MapPriority	Integer	16	5	-		3			
MapTargetID	SCTID	64	18	5				T	
MapRule	String	-	255	-					
MapAdvice	String	-	255	-					

Table A.5: Phrase Search Tables

Table and field names	Type	Size		Partition	Indices				Use of recommended additional indices
		Integer bits	String length		Primary	Recommended		Foreign	
ExcludedWords	Table	-	-	-					
LanguageCode	String	-	8	-	1				
Keyword	Integer	-	8	-	2				
DescWordKey	Table	-	-	-					
Keyword	String	-	8	-	1				
DescriptionID	SCTID	64	18	1	2			D	
ConcWordKey	Table	-	-	-					
Keyword	String	-	8	-	1				
ConceptID	SCTID	64	18	0	2			C	
DescDualKey	Table	-	-	-					
Dualkey	String	-	6	-	1				
DescriptionID	SCTID	64	18	1	2			D	
ConcDualKey	Table	-	-	-					
Dualkey	String	-	6	-	1				
ConceptID	SCTID	64	18	0	2			C	
Word Equivalents	Table	-	-	-		L			
WordBlockNumber	Integer	32	10	-	1				L: To access all instances of a particular word or phrase to allow the relevant blocks to be identified.
WordText	String	-	8	-	2	1			
WordType	Enum	8	2	-		2			
WordRole	Enum	8	2	-					
Duplicate Terms									
See Subset Tables									

Table A.6: Canonical Table

Table and field names	Type	Size		Partition	Indices				Use of recommended additional indices
		Integer bits	String length		Primary	Recommended		Foreign	
Canonical	Table	-	-	-		M	N		
ConceptID1	SCTID	64	18	0	1	1		C	M: To access all or selected canonical Relationships of a specified source Concept. N: To access all or selected canonical Relationships to a specified target Concept.
RelationshipType	SCTID	64	18	0	2	3	2	C	
ConceptID2	SCTID	64	18	0	3		1	C	
RelationshipGroup	Integer	16	5	-	4	2			

Appendix B: Summary of Important Identifiers and Values

B.1 Introduction

This appendix contains tables representing the ConceptIDs applied to Special Concepts that must be recognized for effective implementation of SNOMED CT and the values of enumerated data types used in the distribution tables.

B.2 Important Concept Identifiers

Table B.1: Root Concept and Subtype Relationship

Name	ConceptID	Comments
SNOMED Clinical Terms Concept	138875005	All Concepts are subtype descendants of this Root Concept. The Root Concept has a current Synonym representing the release date.
IS_A	116680003	Relates a Concept to its immediate supertype Concepts.

Table B.2: Top-Level Concepts

Name	ConceptID	Comments
Linkage concept	246061005	
Body structure	123037004	
Situation with explicit context	243796009	
Environment and geographical location	308916002	
Event	272379006	
Clinical finding	404684003	
Observable entity	363787002	
Organism	257495001	
Pharmaceutical / biologic product	373873005	
Physical force	78621006	
Physical object	260787004	
Procedure	71388002	
Qualifier value	362981000	
Record artifact	419891008	
Social context	48176007	
Special concept	123038009	
Specimen	123038009	
Staging and scales	254291000	
Substance	105590001	

Table B.3: Special Concepts

Name	ConceptID	Comments
Special Concept	370115009	A top-level Concept that has as its immediate subtypes a set of Concepts that are used to support the functionality of the terminology rather than to represent real-world Concepts.
Namespace Concept	370136006	A Special Concept with immediate subtypes that each represents a SNOMED CT namespace. There is one Namespace Concept for the SNOMED CT core and one additional Namespace Concept for each Extension for which a namespace-identifier has been allocated.
Navigational Concept	363743006	A Special Concept that has as its immediate subtypes all active Navigation Concepts.
Inactive Concept	362955004	A Special Concept with immediate subtypes that represent each of the possible inactive ConceptStatus values. Each of these has as its immediate subtypes all Inactive Concepts that have that ConceptStatus.

Table B.4: Historical Relationship Type Concepts

Name	ConceptID	Comments
Same as	168666000	Relates a duplicate Concept with an Active Concept that has the same meaning.
Replaced by	370124000	Relates an erroneous Concept to a corrected Concept that replaces it.
May be a	149016008	Relates an ambiguous Concept to the Concepts that represent each of its possible meanings.
Was a	159083000	Relates a Concept to an Inactive Concept that was formerly considered to be one of its supertypes.
Moved to	370125004	Relates to the Namespace Concept for a different namespace in which this Concept is now maintained. The source Concept must have the ConceptStatus “Moved elsewhere” or “Pending move.” The target namespace contains the active version of this Concept with a “Moved from” Relationship to the Concept replaced concept.
Moved from	384598002	Relates to a Concept in another namespace that is replaced by this Concept. The target Concept remains in the original namespace with the inactive ConceptStatus “Moved Elsewhere” or, for a limited period during handover, the active status “Pending Move.” The target Concept has a “Moved to” Relationship pointing to the namespace in which the Concept is now supported.

Table B.5: Valid Relationship Type Concepts – Defining Characteristics[18]

Name	ConceptID	Comments
Access	260507000	
After	255234002	Role Hierarchy – Associated with
Associated finding	246090004	
Associated morphology	116676008	
Associated procedure	363589002	
Associated with	47429007	
Causative agent	246075003	Role Hierarchy – Associated with
Component	246093002	
Clinical course	263502005	
Component	246093002	
Direct device	363699004	Role Hierarchy – Procedure Device
Direct morphology	363700003	Role Hierarchy – Procedure Morphology
Direct substance	363701004	
Due To	42752001	Role Hierarchy – Associated with
Episodicity	246456000	
Finding context	408729009	
Finding informer	419066007	
Finding method	418775008	
Finding site	363698007	
Has active ingredient	127489000	
Has definitional manifestation	363705008	
Has dose form	411116001	
Has focus	363702006	
Has intent	363703001	
Has interpretation	363713009	
Has specimen	116686009	
Indirect device	363710007	Role Hierarchy – Procedure Device
Indirect morphology	363709002	Role Hierarchy – Procedure Morphology
Interprets	363714003	
Laterality	272741003	
Measurement method	370129005	
Method	260686004	
Occurrence	246454002	
Part of	123005000	No longer a defining relationship as of January 2005.
Pathological process	370135005	
Priority	26087009	

[18] Additional attributes are provided in the SNOMED CT attribute hierarchy, but have not applied in SNOMED CT.

Name	ConceptID	Comments
Procedure context	408730004	
Procedure device	405815000	
Procedure morphology	405816004	
Procedure site	363704007	
Procedure site – Direct	405813007	Role Hierarchy – Procedure Site
Procedure site – Indirect	405814001	Role Hierarchy – Procedure Site
Property	37013000	
Recipient category	370131001	
Revision status	246513007	
Route of administration	410675002	
Scale type	370132008	
Severity	246112005	
Specimen procedure	118171006	
Specimen source identity	118170007	
Specimen source morphology	118168003	
Specimen source topography	118169006	
Specimen substance	370133003	
Subject relationship context	408732007	
Surgical approach	424876005	
Temporal context	408731000	
Time aspect	370134009	
Using access device	425391005	Role Hierarchy – Using device
Using device	424226004	Role Hierarchy – Procedure device
Using energy	424244007	
Using substance	424361007	

B.3 Enumerated values

Table B.7: Component Status

Value	Name	Description	Table and Field		
			Concepts ConceptStatus	Descriptions DescriptionStatus	Component History Status
0	Current	Active Component in current use	✓	✓	✓
1	Retired	Withdrawn without a specified reason.	✓	✓	✓
2	Duplicate	Withdrawn from active use because it duplicates another Component.	✓	✓	✓
3	Outdated	Withdrawn from active use because it is no longer recognized as a valid clinical concept/term.	✓	✓	✓
4	Ambiguous	Concept withdrawn from active use because it is inherently ambiguous.	✓		✓
5	Erroneous	Withdrawn from active use as it contains an error. A corrected but otherwise similar Concept has been added to replace it.	✓	✓	✓
6	Limited	Active Concept of limited clinical value as it is based on a classification concept or an administrative definition. Also applies to active Descriptions associated with a limited Concept.	✓	✓	✓
7	Inappropriate	Description withdrawn as this term should not refer to this Concept.		✓	✓
8	Concept Retired	Description contains a valid term used to describe an inactive Concept.		✓	✓
9	Implied	Relationship retired but still true as implied by other Relationships.			✓
10	Moved elsewhere	The Concept is no longer maintained by the organization responsible for the namespace of this ConceptID. The Concept has been moved to another namespace indicated by an associated “Moved to” Relationship.	✓	✓	✓
11	Pending move	The organization responsible for the namespace of this ConceptID has requested that it be moved to another namespace indicated by an associated “Moved to” Relationship. However, the move has not yet been confirmed by the recipient organization. Concepts with this status are still available for active use.	✓	✓	✓

Table B.8: Enumerated Values in Descriptions, Relationships and Subset Tables

Table and Field	Value	Name	Description
Descriptions DescriptionStatus	-	-	See Table B.7
Descriptions Description Type	0	Unspecified	A Description Subset for a language, dialect or realm may assign this as a Preferred Term or Synonym
	1	Preferred	The Preferred Term for the associated Concept.
	2	Synonym	A Synonym for the associated Concept.
	3	FullySpecifiedName	The FullySpecifiedName for the associated Concept.
Relationships CharacteristicType	0	Defining	This Relationship represents a defining characteristic of the source concept. Hierarchical Relationships (e.g. “IS_A”) are also regarded as defining Relationships.
	1	Qualifier	This Relationship represents an optional qualifying characteristic.
	2	Historical	This Relationship relates an inactive Concept to its related active concept.
	3	Additional	This Relationship represents a context-specific characteristic. This is used to convey characteristics of a Concept that apply at a particular time within a particular organization but which are not intrinsic to the Concept.
Relationships Refinability	0	Not refinable	Not refinable.
	1	Optional	May be refined by selecting subtypes.
	2	Mandatory	Must be refined by selecting a subtype.
Subsets SubsetType	1	Language	Descriptions appropriate to a specified language or dialect.
	2	Realm Concept	Concepts appropriate to a specified Realm.
	3	Realm Descriptions	Descriptions appropriate to a specified Realm.
	4	Realm Relationship (for future use)	Relationships appropriate to a specified Realm.
	5	Context Concept	Concepts appropriate to a specified Context Domain.
	6	Context Description	Descriptions appropriate to a specified Context Domain.
	7	Navigation	A set of navigation links used to determine how a Concept hierarchy is arranged and ordered.
	8	Duplicate Terms Subset	A set that includes as its members all Descriptions that share identical terms in a specified Language or Dialect.

Table B.9: Enumerated Values in History and Cross Map Tables

Table and Field	Value	Name	Description
Component History Status	-	-	See Table B.7
Component History Change Type	0	Added	The Component was added to SNOMED CT in this Release Version.
	1	Status Change	The status of the Component has changed since the last Release Version.
	2	Minor Change	A minor change has been made to this Component since the last Release Version. All other changes require the status of the Component.
References Reference Type	1	Replaced by	Refers to a revised replacement for the Component.
	2	Duplicated by	Refers to an identical duplicate for the Component.
	3	Similar to	Refers to a Description that is identical in all respects except for the associated term, which, while not identical, is similar.
	4	Alternative	Refers to one of several alternatives that are similar or equivalent to the Component (e.g. where a single Component is replaced by two or more narrowly defined Components).
	5	Moved To	Refers to the Concept identifying the target Namespace to which a Component has been moved (Status value Moved Elsewhere or is scheduled to be moved (Status value Pending Move).
	6	Moved From	Refers to an original Component in another Namespace which is the source of this current Component.
	7	Refers to Concept	Refers to a Concept which is correctly described by the Term of an inactive Description.
Cross Map Sets MapSetType	0	Unspecified	
	1	Single	All maps are unique one-to-one maps. Each Concept has only one associated Cross Map. Each Cross Map Target contains a single Target Code.
	2	Multiple	Some maps are one-to-many maps but there are no choices. Each Concept has only one associated Cross Map. Some Cross Map Targets contain a list of more than one Target Code.
	3	Choice	Some maps include choices of one-to-one maps but there are no one-to-many maps. Some Concepts have more than one associated Cross Map. Each Cross Map Target contains a single Target Code.
	4	Flexible	Some maps include choices and there are some one-to-many maps. Some Concepts have more than one associated Cross Map. Some Cross Map Targets contain a list of more than one Target Code.

Table B.10: Enumerated Values in Word Equivalents Table

Table and Field	Value	Name	Description
Word Equivalents WordType	0	unspecified	The default value.
	1	word form variant	e.g. “abdomen”, “abdominal”
	2	word equivalents	e.g. “renal”, “kidney”
	3	abbreviation or acronym	e.g. “MI” → “myocardial infarction”
	4	equivalent phrase	e.g. “heart attack”→ “myocardial infarction”
Word Equivalents WordRole	0	unspecified	This default value.
	1	general qualifier	e.g. “mild”, “severe”, “emergency”, etc.
	2	topography	e.g. “arm”, “abdomen”, “abdominal”, etc.
	3	topography qualifier	e.g. “left”, “upper”, etc.
	4	object	including organism or substance
	5	action	e.g. “remove”, “removal” “excision”, etc.
	6	unit of measure	e.g. “ml”, “oz”, etc.

Appendix C: Using Tree View Components for Hierarchy Display

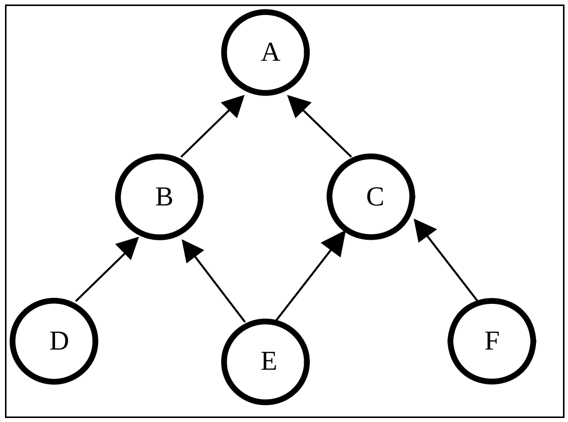
C.1 Introduction

The two examples given below show the creation of a tree view from a small sample hierarchy.

The principals used can be extended to any size or depth of hierarchy.

C.2 Example 1: Show all descendants of Concept "A" in a tree view

ConceptID1	Relationship	ConceptID2
B	IS_A	A
C	IS_A	A
D	IS_A	B
E	IS_A	B
E	IS_A	C
C	IS_A	F



We must process each concept in the hierarchy, starting at 'A'. Add a tree node for 'A', and then query to get the children of 'A'. Process each child recursively, i.e. add a node to the tree view for the child, then query for its children, etc.

Table C.1: Concept to node cross reference

Node	Child Node
1	2
1	5
2	3
2	4
5	4
5	6

Table C.2: Child nodes

Node	ConceptID
1	A
2	B
3	D
4	E
5	C
6	E
7	F

Now we have tree nodes and their children for each Concept. If the nodes have been added to a Windows tree view component, display will be automatic. If a text-based display is being used then the nodes can be output to the screen using the indent style display. Note that the Concept 'E' appears in the tree view twice, under each of its parents.

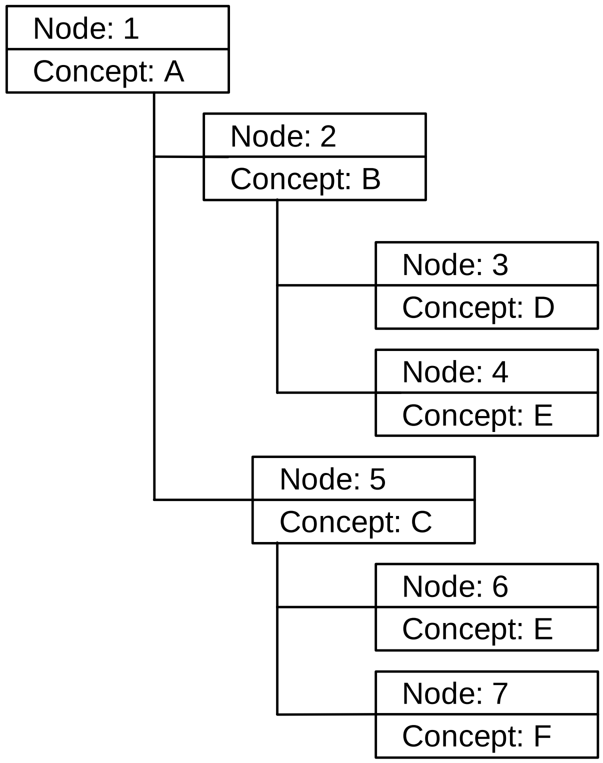


Figure C.1: Tree view of sample hierarchy – descendants of "A"

C.3 Example 2 – Show all ancestors of Concept "E"

In order to construct the tree view, we must start from the top down, so we must create a temporary view of the hierarchy before we can add nodes to the tree view. Query to get the parents of 'E'. Process each parent recursively, i.e. add an entry to the temporary table, stating that 'E' is a child of each of its parents, then query to get its parent, etc. When the top of the tree is reached, a record is kept of the top-level Concept, since this will be the starting point for building the tree view.

Table C.3: Temporary view of the hierarchy

Concept	Child Concept
B	E
C	E
A	B
A	C

We can now use the temporary table information to build the tree view from the top down. Starting at A, add a node to the tree view. Work recursively from the information in the temporary view of the hierarchy to add the descendants of 'A' into the tree view.

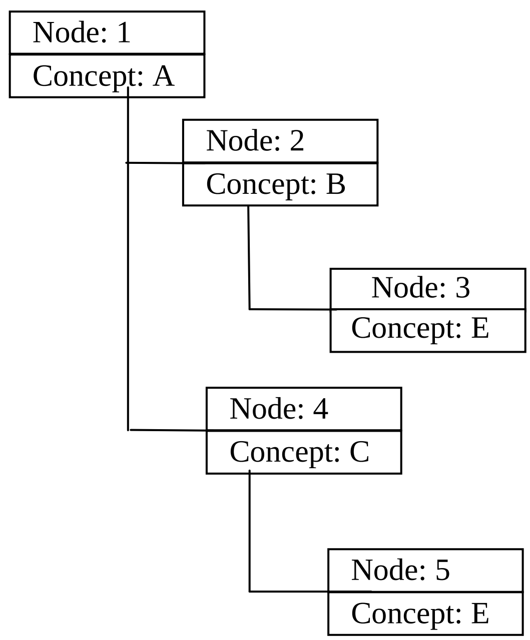


Figure C.2: Tree view of sample hierarchy – ancestors of "E"

Appendix D: SNOMED Clinical Terms and HL7

D.1 Introduction

NOTE:

This section describes an initial view of the interface between SNOMED CT and HL7. Since it was originally drafted, work in the HL7 TermInfo Project has led to increased understanding of the potential gaps and overlaps in semantics that arise at this interface. The HL7 TermInfo Project is developing a document describing the issues and appropriate options that resolve these issues in a way that delivers semantic interoperability. The intention is that when complete that document will be balloted by HL7 and jointly published by the IHTSDO and HL7. Pending publication of the final document, those implementing SNOMED CT and HL7 are recommended to join the HL7 TermInfo group and to refer to the current version of the working documents produced by the project.

HL7 develops standards for the exchange of health related data. The current version of the standard in use is Version 2. Version 3 of the standard is currently being developed, and differs significantly from previous versions. The key advances in HL7 Version 3 are:

- A Reference Information-Model (RIM):
  - This provides a framework for the structure of communicated information.
- A formal development method:
  - This described the various steps to turn a set of requirements into appropriate models and specifications that support communication of the necessary information.
- Separation between structure and content models and implementation technologies:
  - HL7 Version 3 allows alternative implementation based on the same models.
  - Initial implementation use the XML syntax rather than in delimited text files.

This section describes practical issues surrounding the use of SNOMED CT concepts in HL7 Version 3 messages. It assumes that the reader is familiar with HL7 Version 3 – including the HL7 Reference Information Model (RIM) and the Version 3 data types.

In general, it is straightforward to send SNOMED CT concepts in HL7 Version 3 coded fields. HL7 specifies which fields are allowed to contain codes from external terminologies. According to the vocabulary specification for a message, these may be populated with SNOMED CT ConceptIDs.

For each coded field HL7 standards require the specification of a “value-set” that represents the allowable code values in that field. The value-set for a particular RIM attribute vary depending on the actual message specification that uses the RIM attribute. Value-sets may also vary according to realm[19]. Therefore, value-sets need to be defined at the level of individual HL7 specifications with options for variations according to realm.

[19] The HL7 use of the word “realm” usually refers to the sphere of influence of an HL7 International Affiliate or some other body recognized by HL7. For example, HL7-UK is responsible for UK realm specific HL7 vocabularies, messages and guidelines.

An allowable value-set of ConceptIDs can be specified using a Context Concept Subset. This type of Subset allows specification of the field (using the ContextID) and the realm (using the RealmID).

In many cases, there will be more than one way to express the same Concept using SNOMED CT in HL7 messages structures. This is discussed in more detail, but in summary is likely to require:

- Guidelines that recommend one of several possible representations;
- Techniques that determine the equivalence of different representations.

D.2 Overview of HL7V3 concepts

The HL7 Version 3 standards derive their semantic content from the shared HL7 Reference Information Model (RIM), and are implemented in Extensible Markup Language (XML).

All Version 3 products derive their semantic content from the RIM. The RIM is general and expressive enough to serve as the backbone for all HL7 Version 3 specifications, but is too general to be used directly by any specification. Intermediary models are used to constrain the RIM for use in a particular specification. For this reason, SNOMED CT will typically apply these constructs to these intermediary models, rather than to RIM itself. The first intermediary model is known as Domain Message Information Model (D-MIM), which constrains the portion of the RIM used by a committee in the derivation of all their messages. This model is further constrained to a Refined Information Model (R-MIM), and then to a Hierarchical Message Definition (HMD), which specifies a set of Message Types. Thus, there will be several D-MIMs derived from the RIM; there will be several R-MIMs derived from each D-MIM; and there will be several HMDs derived from each R-MIM.

HL7 also defines Common Message Element Types (CMET) – which are RIM fragments that can be used by a number of D-MIMs. A CMET is expressed as an R-MIM with derived HMDs and Message Types, which can then be re-used in several different messages.

The RIM includes a new set of data types developed for use within the HL7 Version 3 family of standards. These data types include some of the familiar ones used in HL7 V2.x messaging, such as STRING (ST), INTEGER (INT), and TIME STAMP (TS); as well as new data types such as ENCODED DATA (ED), which supports multi-media; and INTERVAL of TIME (IVL<TS>), which allows for the expression of a time range. Data types that can carry ConceptIDs include: Coded with Equivalents (CE), which carries a code, the name of the coding scheme the code is drawn from, and a display name corresponding to the code; and allows synonyms to be transmitted – such as an HL7 code and its equivalent SNOMED code; and Concept Descriptor (CD), which builds on the CE by supporting the post-coordination of codes (or, stated in another way, the combining of codes from a terminology to create a new concept). The CD data type is described in greater detail below, along with implementation guidance for using it to compose SNOMED concepts.

- Each RIM attribute has a specified data type. RIM attributes of type CE or CD can also have a specified vocabulary domain. These domains can include HL7-defined concepts or can be drawn from HL7-recognized coding systems such as LOINC or SNOMED CT. Vocabulary domains have a coding strength that can be “Coded, No Extensions” (CNE), in which case the only allowable values for the field are those in the vocabulary domain; or “Coded, With Extensions” (CWE), in which case values other than those in the vocabulary domain (such as local codes) can be used if necessary.

- The vocabulary domain specifications stated in the RIM always refer to a complete vocabulary domain. That is, at the RIM level there is no specialization based on realm of use or on the context and needs of a specific message. As RIM attributes are specialized to suit a specific message context, the domain of the attribute can be reduced (constrained) to reflect the specialization. A vocabulary domain that has been constrained to a particular realm and coding scheme (such as SNOMED CT) is called a "value set."

To summarize the relationship between vocabulary and HL7 specifications:

- At the level of the RIM, attributes of type CE or CD will declare a single "vocabulary domain," which can be CWE or CNE (but typically will be CWE).
- An individual message specification can further restrict this vocabulary domain.
- A vocabulary domain can be constrained to a particular realm and coding scheme, creating a "value set."
- Value sets will be SNOMED CT subsets, appropriately specified to the particular realm of use. (In other words, the exact SNOMED CT subset for a particular field in a particular message can differ between the United Kingdom and the United States, although both subsets must be a constraint on the RIM-level vocabulary domain).

D.3 Retrieval and aggregation of HL7 representations of SNOMED CT encoded data

D.3.1 Overview

Because concepts can be represented in more than one way, there needs to be an approach to aggregate the various representations. The process relies on an understanding of both HL7 Version 3 and SNOMED CT, and involves a mapping between the two so that concepts sent in HL7 V3 can be converted into a standard form prior to aggregation.

It is useful to summarize the steps at a high level. These steps will then be further described in greater detail in the sections that follow:

- Retrieve an HL7 V3 message
    - A receiver of one or more HL7 messages will need to be able to extract the coded information in the message and aggregate it with concepts sent in other messages or with concepts stored in a data repository.
    - The entire discussion of aggregation in this section assumes that valid and conformant HL7 messages are received.
- Syntactic transformation
    - Convert the concepts in the message into a "canonical form" (using a derived equivalence between HL7 RIM attributes and SNOMED CT relationship types).
    - The use of guidelines and templates can constrain the inherent flexibility of an HL7 message, and can decrease the number of sub-steps required to perform a canonical transformation. Tightly coupled systems can take this into account, and establish bilateral agreements that will minimize message variability.
- Aggregation
    - Aggregate the various representations, all expressed in a common canonical form.

- Use techniques that query for the primary code AND the semantic properties of the concepts of interest.

D.3.2 Syntactic transformation

Aggregation of concepts represented in more than one way is made easier by mapping them into a common form. The common form described here is known as the long canonical form (further described in Spackman [01Sp]), which maximally decomposes concepts into their most proximate primitive supertypes, and their most specific values for each defining relationship. As noted above, this requires there to be some combination of mapping rules and guidelines.

Again, the objective in this step, which is required prior to aggregation, is to transform HL7 instances into a common form. Examples of SNOMED CT concepts decomposed in to their long canonical form:

Concept	Defining Attribute	Attribute Value
Hypophysectomy 52699005	IS_A	Procedure 71388002
Hypophysectomy 52699005	METHOD	excision-action 129304002
Hypophysectomy 52699005	PROCEDURE-SITE	Pituitary structure 56329008

Concept	Defining Attribute	Attribute Value
Brain surgery (P1-91010)	IS_A	Procedure 71388002
Brain surgery (P1-91010)	METHOD	surgical-action 129284003
Brain surgery (P1-91010)	PROCEDURE-SITE	brain structure 12738006

Concept	Defining Attribute	Attribute Value
Repair of incarcerated inguinal hernia 12646000	IS_A	Procedure 71388002
Repair of incarcerated inguinal hernia 12646000	METHOD	surgical-repair-action 129342008
Repair of incarcerated inguinal hernia 12646000	DIRECT-MORPHOLOGY	incarcerated-hernia 110418002
Repair of incarcerated inguinal hernia 12646000	PROCEDURE-SITE	inguinal-canal-structure 90785001

The syntactic transformation from HL7 instances to the long canonical form can be performed by following these steps:

STEP: Adopt SNOMED CT value sets.

- The Context Concept Subset provides an effective way to define value sets for coded fields in HL7 specifications. Separate Subsets can be defined for each context (field or attribute) in each realm (country or other jurisdiction). As deemed appropriate, the IHTSDO may either release these Subsets or enable other authorized organizations to do so. Use of agreed value-sets is highly recommended.

STEP: Understand the various ways concepts can be represented in HL7 V3 specifications.

- Users of SNOMED CT in HL7 will need to understand the various ways concepts can be represented in HL7 V3 specifications in order to understand the variations needing to be anticipated.

STEP: Mapping from HL7 to SNOMED CT.

- Given the stated value sets and given an understanding of the various HL7 V3 specifications, a mapping between an HL7 V3 message and its corresponding representation in SNOMED CT long canonical form can be performed.

Suggested SNOMED CT value sets

Introduction

The IHTSDO recognizes that there is a requirement for appropriately maintained value-sets for use in the coded attributes of particular message types. Detailed plans for the development and maintenance of these value-sets have not yet been agreed. However, it is anticipated that this task will involve cooperation between HL7 and the IHTSDO.

A few **draft** value sets are presented here to illustrate that many of the most commonly required sets can be specified as subtype descendants of a single Concept.

As described above, each RIM attribute of type CE or CD has an HL7-declared vocabulary domain. This vocabulary domain can be constrained to specific SNOMED CT value sets. In the process of deriving an HL7 specification from the RIM, the R-MIM or HMD can apply tighter constraints than those at the RIM level. SNOMED value sets are defined to the level of individual message and document specifications.

RIM value sets

The following value sets are declared at the RIM level. Downstream intermediary models and specifications can apply further constraints.

Table D.1: Concept representation in HL7 V3 specification

RIM Attribute	SNOMED Value Set
Access.target_site_cd	Descendant of Body structure 123037004
Non_person_living_subject.taxonomic_classification_cd	Descendant of Living organism 4748007
Observation.target_site_cd	Descendant of Body structure 123037004
Procedure.approach_site_cd	Descendant of Procedural approach 103379005
Procedure.method_cd	Descendant of Action 129264002
Procedure.target_site_cd	Descendant of Body structure 123037004

In general, it is straightforward to send SNOMED CT concepts in HL7 Version 3 coded fields. HL7 specifies which fields are allowed to contain codes from external terminologies. According to the vocabulary specification for a message, these may be populated with SNOMED CT ConceptIDs. However, there is potential to deviate from this basic pattern, given the flexibility of HL7 specifications and this warrants a special discussion of the various ways ConceptIDs (or equivalent concepts represented without the use of ConceptIDs) can be transmitted in HL7 V3.

RIM attributes

ConceptIDs are only sent in RIM attributes of data type CE or CD that are declared as having a CWE vocabulary domain. If HL7 has declared a value set for a particular coded attribute (for use within a particular realm), their policy for extending this value set with ConceptIDs are:

- User must send a value from the HL7-declared value set if a valid value exists.
- If the user wants to also send a synonymous ConceptID, it comes after the HL7 value.

- If there is no suitable value in the HL7-declared value set, the user sends just the ConceptID.

RIM attributes of data type CS (Code Simple) usually have an HL7-defined CNE vocabulary domain, and thus cannot transmit ConceptIDs. While ConceptIDs cannot be sent in these fields, the HL7 concepts that are sent will often have overlapping representations in SNOMED CT.

Non-SNOMED vocabulary

RIM attributes of data type CE or CD can transmit codes from any HL7-registered coding scheme. The particular message may suggest the need for a particular coding scheme – such as ICD codes for regulatory reporting, or Laboratory LOINC codes for lab observations.

SNOMED CT guidelines for dealing with multiple coding schemes include:

- The use of multiple coding schemes is strongly discouraged.
- The allowable set of coding schemes needs to be constrained by trading partners.
- The IHTSDO has mapped several common coding schemes (such as ICD and integrated with Laboratory LOINC) into SNOMED. Where it is necessary to use non-ConceptIDs, the strong preference is to draw codes from coding schemes that have been mapped to SNOMED CT. If the "mapping category" subcomponent of the CE data type conflicts with the mapping category defined by SNOMED CT the SNOMED CT representation will prevail for SNOMED CT encoded attributes.
- RIM attributes of data type CS with a CWE domain should ONLY be extended with ConceptIDs. The use of codes from any other coding scheme is strongly discouraged.
- The use of any other coding scheme within the CD data type is strongly discouraged. (Note that in some circumstances, such as the use of HCPCS modifiers, there will be a need to put non-ConceptIDs into the CD data type). The use of more than one coding scheme within a single instance of the CD data type is prohibited.

Data types

RIM attributes of data type CE or CD can transmit ConceptIDs. Additionally, some RIM attributes are defined as having a "set" of CE or CD values. When a RIM attribute carries a set of values: "The elements are contained in no particular ordering. All elements in the set are distinct; the same element value cannot be contained more than once in the set."

Each of these data types has some unique considerations:

- The CE data type:
  - The CE data type contains a "mapping category" component that shows the type of mapping (e.g. synonyms, broader-to-narrower) between the primary code and the translated code. SNOMED CT mappings are to be used over a value sent for mapping category if the two values conflict.
- The CD data type:
  - The Concept Descriptor data type contains a grammar for the post-coordination of ConceptIDs. The post-coordination grammar allows you to assign modifiers (roles and their values, where values can themselves be further modified) to a concept. For instance, the following concept can be constructed in a CD data type: (Hypophysectomy (approach=Transfrontal approach))

The Concept Descriptor data type

Given the expressive power of the Concept Descriptor (CD) data type, several aspects of its use warrant further discussion:

- **SNOMED CT Qualifiers:** SNOMED CT Qualifiers should be expressed in the CD data type. Where SNOMED CT Qualifiers are defined, their use is highly recommended. The preferred way of instantiating SNOMED CT Qualifiers in HL7 V3 is within the CD data type.
- **Allowable roles:** SNOMED CT defining attributes should be used preferentially as roles in the CD data type. Also, for each defining attribute, SNOMED specifies the type of concept where it is allowable. For instance, the BRANCH-OF defining attribute can only be used in the definition of topographic concepts. Additional allowable roles include descendents of Linkage concept 106237007. These additional roles should only be used when and if it is necessary to transmit semantics that cannot be expressed using RIM structures, qualifiers, or defining attributes.
- **Nested role-value pairs:** In general, nested role-value pairs are analogous to SNOMED role groupings.

  Example:

  - Liver-cancer-with-lung-metastases (assoc-morphology :: cancer (finding-site=liver)) (assoc-morphology=metastases (finding-site=lung))

  is equivalent to

  - Liver-cancer-with-lung-metastases ((assoc-morphology=cancer) (finding-site=liver)) ((assoc-morphology=metastases) (finding-site=lung)).

  Note that only those defining attributes that can be applied to the primary concept should be included in the CD. For instance, the following would be an **invalid** post-coordination expression, because it uses the “branch-of” role in the post-coordinated definition of a procedure concept:

  - Excision-of-artery-X (procedure-site=artery-x (branch-of=artery-y)).
- **Where to post-coordinate:** It is possible to post-coordinate in more than one way because there is overlap between SNOMED CT defining characteristics and HL7 RIM attributes (e.g. the defining characteristic “procedure-site” and the HL7 attribute “Procedure.target_site_cd”). HL7’s policy on this is:
  - A user does not need to break apart a pre-coordinated concept in order to maximally populate applicable RIM attributes. (For instance, if the procedure site is pre-coordinated in the procedure concept, the user does not need to redundantly populate the Procedure.target_site_cd attribute).
  - Applicable RIM attributes should be used where possible. Users can choose to redundantly post-coordinate in the CD data type. (For instance, if the procedure site is post-coordinated in the originating record, it should be conveyed in the HL7 Procedure.target_site_cd attribute, and can optionally post-coordinate the same body site in the procedure.cd attribute).
- SET of CE or CD
  - All elements in a set are considered logically AND’d to one another – just as if there were two of the same defining characteristics in a pre-coordinated concept

definition. For example, this HL7 instance: (Excision-of-artery-of-head (Procedure.target_site_cd = “artery,head”)) is analogous to this:

- Excision-of-artery-of-head (procedure-site=artery) (procedure-site=head)).

Mapping from HL7 V3 to SNOMED

Note:

Only a few draft mappings are currently presented. These will be refined after HL7 V3 closes ballot.

As noted above, given the stated value sets and given an understanding of the various HL7 V3 specifications, a transformation from an HL7 V3 instance in to the long canonical form can be performed.

CMET to SNOMED mappings

The following mappings are declared at the CMET level, and are applicable wherever the particular CMET is used.

Clinical_Procedure CMET

Detailed mapping:

CMET Attribute (data type)	SNOMED defining attribute	SNOMED Value Set	Constraints / Comments
cd (CD)	IS_A	Descendent of Procedure 71388002	
approach_site_cd (Set<CD>)	APPROACH	Descendent of Procedural approach 103379005	Constrain Set<CD> to Set<CE> because SNOMED CT does not have sets of defining attributes for approach.
target_site_cd (Set<CD>)	PROCEDURE-SITE	Descendent of Body-structure 123037004	
method_cd (Set<CE>)	METHOD	Descendent of Action 129264002	

Example of HL7 instance:

<Observation>

<!-- other elements here -->

<Clinical_procedure>

<cd code="**807005**"

displayName="**Brain excision**" codeSystem="**SNOMED CT**"/>

<approach_site_cd code="**65519007**"

displayName="**Transfrontal approach**" codeSystem="**SNOMED CT**"/>

<target_site_cd code="**56329008**"

displayName="**Pituitary structure**" codeSystem="**SNOMED CT**"/>

</Clinical_procedure>

</Observation>

Mapped into SNOMED CT:

Concept	Defining Attribute	Attribute Value
HL7-instance	IS_A	Brain excision 807005
HL7-instance	APPROACH	transfrontal-approach 65519007
HL7-instance	PROCEDURE SITE	Pituitary structure 56329008

Transformed into SNOMED CT long canonical form*:

Concept	Defining Attribute	Attribute Value
HL7-instance	IS_A	Procedure 71388002
HL7-instance	APPROACH	transfrontal-approach 65519007
HL7-instance	METHOD	excision-action 129304002
HL7-instance	PROCEDURE SITE	Pituitary structure 56329008

*Transformation into long canonical form uses the Relationships Table and the Canonical Table in the manner outlined in section [6.7.2].

R-MIM to SNOMED mappings

The following mappings are declared at the R-MIM level, and are applicable wherever the particular CMET is used.

Substance administration R-MIM

Detailed mapping:

Attribute (data type) <= Vocab Domain	SNOMED defining attribute	SNOMED Value Set	Constraints / Comments
class_cd (CS) <= “SBADM”	IS_A	Administration 108234007	Recognize that HL7 "substance administration acts" as a kind of SNOMED (drug) administration procedure.
cd (CD)	DIRECT-SUBSTANCE	Descendent of Substance 105590001	Constrain CD to CE. *NOTE: The HL7 specification transmits the substance code in this field.*
route_cd (CD) <= MedAdministrationRoute	METHOD; APPROACH	METHOD: Descendent of Administration action 129445006 APPROACH: Descendent of Route of Administration 103389009	Constrain CD to CE. *NOTE: The HL7 administration route is a pre-coordinated mix of the action (e.g. injection vs. infusion) along with the route (e.g. oral vs. intravenous).*

HL7 V3 Route of Administration vocabulary domain	Mapped SNOMED Representation
IAINJ Intraarterial injection	METHOD Injection-action 129326001 APPROACH Intra-arterial route 58100008
IM Intramuscular injection	METHOD Injection-action 129326001 APPROACH Intramuscular route 78421000
IVINJ Intravenous injection	METHOD Injection-action 129326001 APPROACH Intravenous route 47625008
NASINHL Nasal Inhalation	METHOD Inhalation-action 112239003 APPROACH Nasal route 46713006
TRNSDERM Transdermal Application	METHOD Application-action 129425003 APPROACH Transdermal route 45890007

Example of HL7 instance:

Intravenous injection of cortisone:

<Act>

<Act.class_cd code="**SBADM**" displayName="**substance administration**"/>

<Act.cdcode="**111150008**" codeSystem="**SNOMEDCT**" displayName="**cortisone**"/>

<Act.route_cd code="**43060002**" codeSystem="**SNOMED CT**"

displayName="**Intravenous injection**"/>

</Act>

Mapped into SNOMED CT:

Concept	Defining Attribute	Attribute Value
HL7-instance	IS_A	Administration 108234007
HL7-instance	DIRECT SUBSTANCE	Cortisone 111150008
HL7-instance	METHOD	injection-action 129326001
HL7-instance	APPROACH	transvenous approach 103386002

Transformed into SNOMED CT long canonical form*:

Concept	Defining Attribute	Attribute Value
HL7-instance	IS_A	Procedure 71388002
HL7-instance	DIRECT SUBSTANCE	Cortisone 111150008
HL7-instance	METHOD	injection-action 129326001
HL7-instance	APPROACH	transvenous approach 103386002

*Transformation into long canonical form uses the Relationships Table and the Canonical Table in the manner outlined in section [6.7.2].

D.3.3 *A*ggregation

What has become clear is that information retrieval is enhanced by aggregation techniques that query for the primary code AND the semantic properties of the concepts of interest. The technique described here is further described by Brown, et al [00Br].

To facilitate the discussion of aggregation, this subsection now assumes that all Concepts to be aggregated (Concepts received in HL7 messages, Concepts in patient records, Concepts in SNOMED CT) exist or can be transformed into the long canonical form. Steps involved in aggregation include:

STEP: All *Concepts* to be aggregated can be transformed into the long canonical form.

- This material is covered above, and is assumed for the rest of the discussion.

STEP: Think of "aggregation" as "classification".

- Think of the query as itself being a Concept (as opposed to, say, a SQL statement), and think of the results as those Concepts that would classify as descendants of the query concept. The query concept is itself expressed in the long canonical form to facilitate classification.

Example:

Assume the goal is to search for all patients who have had an “excision of the pituitary gland” or to find all patients who have had an “operation on the pituitary gland.” To approach this example, we will

1. Receive some HL7 messages and convert them into long canonical form;
2. Express the query “excision of the pituitary gland” and “operation on the pituitary gland” in the long canonical form; and
3. Determine which of the HL7 instances meet the criteria of the query.

Receive some HL7 messages and convert them into SNOMED CT long canonical form

Here are some sample HL7 instances in native form:

```
<Observation>
 <other-stuff/>
 <Clinical_procedure>
  <cd code="52699005" displayName="Hypophysectomy" codeSystem="SNOMED CT"/>
  <txt>Hypophysectomy by transfrontal approach</txt>
  <approach_site_cd code="65519007"
   displayName="Transfrontal approach" codeSystem="SNOMED CT"/>
 </Clinical_procedure>
</Observation>
<Observation>
 <other-stuff/>
 <Clinical_procedure>
  <cd code="42699003" displayName="Brain incision" codeSystem="SNOMED CT">
   <modifier code=”49755003” displayName=”lesion” codeSystem="SNOMED CT">
    <name code=”363700003” displayName=”DIRECT-MORPHOLOGY” codeSystem="SNOMED CT"/>
   </modifier>
  </cd>
  <txt>Incision of lesion of posterior lobe of pituitary gland</txt>
  <target_site_cd code="37512009"
   displayName="Pituitary posterior lobe" codeSystem="SNOMED CT"/>
 </Clinical_procedure>
</Observation>
<Observation>
 <other-stuff/>
 <Clinical_procedure>
  <cd code="86146009"
   displayName="Partial excision of pituitary gland by transfrontal approach"
   codeSystem="SNOMED CT"/>
  <txt>
   Partial excision of pituitary gland by transfrontal approach
  </txt>
  </Clinical_procedure>
</Observation>
```

Mapped into SNOMED:

Concept	Defining Attribute	Attribute Value
HL7-instance-1	IS_A	Hypophysectomy 52699005
HL7-instance-1	APPROACH	transfrontal-approach 65519007
HL7-instance-2	IS_A	Brain incision 42699003
HL7-instance-2	DIRECT MORPHOLOGY	Lesion 49755003
HL7-instance-2	PROCEDURE SITE	pituitary-posterior-lobe 37512009
HL7-instance-3	IS_A	Partial excision of pituitary gland by transfrontal approach 86146009

Transformed into SNOMED long canonical form*:

Concept	Defining Attribute	Attribute Value
HL7-instance-1	IS_A	Procedure 71388002
HL7-instance-1	APPROACH	transfrontal-approach 65519007
HL7-instance-1	METHOD	excision-action 129304002
HL7-instance-1	PROCEDURE SITE	Pituitary structure 56329008
HL7-instance-2	IS_A	Procedure 71388002
HL7-instance-2	METHOD	incision-action 129287005
HL7-instance-2	DIRECT MORPHOLOGY	Lesion 49755003
HL7-instance-2	PROCEDURE SITE	pituitary-posterior-lobe 37512009
HL7-instance-3	IS_A	Partial-excision 38829003
HL7-instance-3	APPROACH	transfrontal-approach 65519007
HL7-instance-3	METHOD	excision-action 129304002
HL7-instance-3	PROCEDURE SITE	Pituitary structure 56329008

*Transformation into long canonical form uses the Relationships Table and the Canonical Table in the manner outlined in section [6.7.2].

Express the query "excision of the pituitary gland" and "operation on the pituitary" in long canonical form

The long canonical form of "excision of the pituitary gland" is:

Concept	Defining Attribute	Attribute Value
excision of pituitary gland	IS_A	procedure 71388002
excision of pituitary gland	METHOD	Excision-action 129304002
excision of pituitary gland	PROCEDURE SITE	pituitary structure 56329008

The SNOMED long canonical form of "operation on the pituitary gland" is:

Concept	Defining Attribute	Attribute Value
excision of pituitary gland	IS_A	procedure 71388002
excision of pituitary gland	METHOD	surgical-action 129284003
excision of pituitary gland	PROCEDURE SITE	pituitary structure 56329008

Determine which of the HL7 instances meet the criteria of the query

The aggregation process now checks to see which of the concepts in long canonical form would classify as a descendent of the query concept.

Concepts that aggregate under "excision of the pituitary gland" would be those procedures matching ALL the following criteria:

- At least one supertype is a kind of *Procedure* 71388002.
- At least one METHOD is a kind of *excision-action* 129304002.
- At least one PROCEDURE SITE is a kind of *Pituitary structure* 56329008.

Concepts that aggregate under "operation on the pituitary gland" would be those procedures matching ALL of the following criteria:

- At least one supertype is a kind of *Procedure* 71388002.
- At least one METHOD is a kind of *surgical-action* 129284003.
- At least one PROCEDURE SITE is a kind of *Pituitary structure* 56329008.

Given this, HL7-instance-1 and HL7-instance-3 would aggregate under "excision of the pituitary gland," and all of the HL7 instances would aggregate under "operation on the pituitary." In addition, patient records containing the pre-coordinated SNOMED concept "*Total excision of pituitary gland by transfrontal approach* 67749007" would aggregate under both queries, whereas "*Incision of pituitary gland* 6846004" would only aggregate under the second query "operation on the pituitary gland."

D.4 Open issues

D.4.1 Observation.cd vs. Observation.value

When sending HL7 observations, there is some flexibility when determining the *Concept* values that populate the slot that holds the code for the type of observation (Observation.cd) and the slot that holds the observation value (Observation.value). Here is Stan Huff's classic example, using LOINC codes in HL7 V2.x messages:

Interface A (Rheumatologist view)

OBX|1|CE|4821-5^HLA-B27^LOINC|1|G-A203^Present^SNI|...

Interface B (Paternity testing)

OBX|1|CE|4694-6^HLA-TYPE^LOINC|1|F-C4327^B27^SNI|...

Draft HL7 and SNOMED CT guidelines currently recommend the following:

- When exchanging information about observations:
  - Observation.cd can be populated with a descendant of *Observation procedure* 122544007.
  - Observation.cd can be populated with a descendent of Observable entity 363787002.

Observable entity values semantically represent the measured component of an observation procedure. There when they are used, they should be assumed to be an observation of the particular observable entity. For example, you can send the observation procedure code "*total body water measurement* 241419008", or, if you can send the observation entity code "*total body water* 251837008," you would assume it is an observation of total body water, and therefore logically equivalent to the observation procedure code.

- When exchanging information about findings and diseases:
  - Observation.cd can be populated with a descendent of Qualifier for *type of diagnosis* 106229004.
  - Observation.value should be populated with descendents of *Finding* 246188002 or descendents of *Disease* 64572001.

D.4.2 Aggregation limitations

Note that the above algorithms would aggregate a Concept under a query concept whether the Concept transmitted in the HL7 instance is, say, stated to be absent or slated to be present in a family member rather than in the patient. The HL7 RIM contains formalisms for context and inheritance (e.g. Act_relationship.context_control_cd), and for the explicit negation of findings (e.g. Act.negation_ind). It is strongly recommended that these formalisms be considered when extracting data from HL7 instances. The exact process by which these formalisms should be used need further study.

Note also that the above algorithms are based on the shared use of definitional attributes among Concepts that are all expressed in the long canonical form. Where SNOMED CT does not include a sufficiently fine-grained defining characteristic (e.g. a definitional attribute to describe the color in "red rash"), it is not possible to directly aggregate all Concepts with red findings.

D.4.3 Aggregation using a classifier

The process for aggregation of pre- and post-coordinated SNOMED concepts relies on a syntactic transformation into a standard form. The sections above used the SNOMED long canonical form as the standard form, but there are other standard forms – such as description logic formalism. One advantage to using a description logic formalism as the standard form is that classification algorithms (including the same ones used to classify SNOMED CT) can perform additional aggregation, above and beyond that described above – such as right-identity mapping, and classifications based on role groupings.

It's unclear at this time just what role of the classifier should be in data aggregation. Perhaps it should serve as the gold standard for completeness, by which other techniques are to be compared.

D.4.4 Expressing laterality

The SNOMED CT LATERALITY 272741003 attribute is used in the pre-coordinated definition of *Body structure* 123037004 concepts. In post-coordinated expressions, it can potentially be used to modify body structures, procedures, diseases, and findings.

When laterality is used to modify procedure, disease or finding concepts that refer to a single body structure, it is actually modifying that underlying body site within the main concept.

LATERALITY can also be used to modify procedure, clinical finding or disease concepts that have more than one body site, but only if both sites have the same value for laterality (both left, both right, or both bilateral). If possible, when composing (post-coordinating) a Concept that involves multiple sites, with differing values for laterality, the record structure should record these as distinct Concepts (one concept for each site). If users chose to compose a single

Concept that involves multiple sites, with differing values for laterality, then they must either: 1) Use existing pre-coordinated SNOMED body structure concepts that represent laterality (e.g. Structure of right ankle); or, 2) Represent LATERALITY as a qualifier for each body structure within the main Concept, as applying laterality to the entire Concept would not be appropriate. This second option may be difficult to implement.

D.5 References

- [RIM] HL7 Reference Information Model (*www.hl7.org/library/data-model/RIM/modelpage_non.htm*)
- [98Be] Beeler GW Jr. HL7 version 3 – an object-oriented methodology for collaborative standards development. Int J Med Inf. 1998 Feb; 48(1-3): 151-61.
- [00Ba] Bakken S, Campbell KE, Cimino JJ, Huff SM, Hammond WE. Toward vocabulary domain specifications for health level 7-coded data elements. JAMIA 2000; 7(4): 333-42.
- [01Sp] Spackman, KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Fall AMIA 2001 (in press).
- [00Br] Brown PJB, Sonksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. JAMIA 2000; 7:391-403.
- [98Do] Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a “Lexically Assign, Logically Refine” strategy for semi-automated integration of overlapping terminologies. JAMIA 1998; 5(2): 203-13.

Appendix E: Developer Toolkit

E.1 Introduction

Several components of SNOMED® Clinical Terms have been collected into a Developer Toolkit to assist software application developers. The Developer Toolkit contains the actual files that support each SNOMED CT release.

E.2 Components of the Developer Toolkit

The components of the Developer Toolkit include:

Category	Component	Comments
Indexes	Excluded Words Table Description Word Key Table Concept Word Key Table Description Dual Key Table Concept Dual Key Table	These indexes have been pre-generated using the SNOMED CT Descriptions Table (English Language).
Navigation Hierarchies	Two samples provided: • SNOMED CT top-level hierarchies • CTV3	
Additional Components	Duplicate Terms Table Word Equivalents Table	The Duplicate Terms Table uses the subset mechanism. An extended version of the Toolkit includes the programs to generate the search indexes.

Appendix F: How the Short Canonical Form Is Derived

F.1 Introduction

This section is based on the process for generating the short canonical form used to generate the "canonical table" included in this release. This version of the short canonical form, while still useful, is no longer regarded as the best form for use in testing subsumption of pre and post-coordinated expressions.

Two draft documents covering this topic in greater detail are available from the IHTSDO. The first of these discusses the various abstract models and representational forms that may be applicable to SNOMED CT, and the second provides detailed advice on transformation to common normal forms in which alternative representations can be readily compared.

These additional documents are currently in draft form for discussion within the IHTSDO. However, they are recommended to implementers seeking a more complete, detailed and up to date understanding of alternative forms, transformations and comparisons.

This section describes the process by which the short canonical form is derived from the Relationships and Concepts that are included in a SNOMED CT release. This process is documented to enable developers using this data to understand its relationship with data released in the core tables.

F.2 Overview

Before the SNOMED CT core tables are released, they are processed using a description logic classifier. Description logic classification ensures that the Relationships in releases are logically consistent. This is an essential precursor to generation of the short canonical form.

Working from this logically consistent baseline, each Concept is tested as follows:

- Identify its proximal Primitive supertype Concept(s)
  - For each of these add an "IS_A" subtype Relationship to the Canonical Table.
- Identify its unshared defining characteristics
  - For each of these add an appropriate Relationship to the Canonical Table.

The Concept being tested is referred to in the following more detailed descriptions as the FocusConcept*.*

F.3 Description logic classification

Description logic classification is undertaken prior to each release of SNOMED CT to produce a logically consistent se of Relationships. This process ensures that:

- The set of "IS_A" subtype Relationships:
  - Is complete based on the current stated defining characteristics.
    - Additional inferred "IS_A" Relationships are automatically added where necessary.
  - Contains no cyclical definitions.
    - Cyclical definitions, such as ***A*** IS_A ***B*** + ***B*** IS_A ***C*** + ***C*** IS_A ***A*** are automatically detected and manually resolved.
  - Refer only to most proximate supertypes.

    - ° Stated Relationship sets such as ***A*** IS_A ***B*** + ***B*** IS_A ***C*** + ***A*** IS_A ***C*** are rationalized by excluding the redundant distant definition ***A*** IS_A ***C***.
- Descendant subtypes inherit appropriate defining characteristics.
    - If it is stated that ***A*** IS_A ***B*** + ***B*** has property ***C***, it must also be true that either:
        - ° ***A*** has property ***C***

      or

        - ° ***A*** has property ***D*** and ***D*** IS_A ***C***

      If neither of these is stated, the inferred *Relationship* ***A*** has property ***C*** is added.

F.4 Identifying proximal primitive supertypes

1. Select all rows in the Relationships Table where:
    - ConceptId1 refers to the FocusConcept AND
    - RelationshipType refers to the "IS_A" Concept.
2. For each of the Reference Concepts pointed to by ConceptId2 in each of the selected rows:
    - If IsPrimitive =0:
        - Select all rows in *Relationships* Table where
            - ConceptId1 refers to the ReferencedConcept AND
            - RelationshipType refers to the "IS_A" Concept.
        - For each of these selected rows repeat step 2 (recursive).
    - If IsPrimitive =1:
        - Add ReferenceConcept to the list of PrimitiveSupertypes.
3. If in recursion as a result of (2b):
    - Exit that level of recursion and continue processing the rows selected at previous level.
4. For each Concept in the list of PrimitiveSupertypes:
    - If any of the other PrimitiveSupertypes have this Concept as a PrimitiveSupertype
        - Remove the Concept from the list of PrimitiveSupertypes
5. For each Concept in the list of PrimitiveSupertypes:
    - Create a Relationship in the Canonical Table in which
        - ConceptId1 refers to the FocusConcept
        - RelationshipType refers to the "IS_A" Concept
        - ConceptId2 refers to the PrimitiveSupertype
        - Relationship group = 0

F.5 Identifying unshared defining characteristics

1. Select all rows in the Relationships Table where:
    - ConceptId1 refers to the FocusConcept AND
    - RelationshipType does not refer to the "IS_A" Concept AND
    - CharacteristicType = 0 (Defining characteristic)
2. Mark each of the selected rows as a provisional row for the Canonical Table.
3. For each Concept in the list of PrimitiveSupertypes:
    - Select all rows in the Relationships Table where:
        - ConceptId1 refers to the PrimitiveSupertype AND
        - RelationshipType does not refer to the "IS_A" Concept AND

- CharacteristicType = 0 (Defining characteristic)
- For each selected row:
  - Unmark any provisional row that matches the selected row on the following attributes:
    - RelationshipType AND ConceptId2

4. Add each remaining marked provisional row to the Canonical Table.
   - The fields are populated with values of the corresponding fields in the Relationships *Table*.

F.6 Illustration of the short canonical form

F.6.1 Introduction

This example describes the differences between a full set of Relationships (Figure F.1) and the short canonical form (Figure F.2). In Figure F.2:

- Each box in these diagrams represent a Concept:
  - The text in the lower part of the box represents the defining characteristics of the Concept.
  - Shaded boxes containing the word "Primitive" indicate that a Concept is Primitive – that is, the concept is not fully defined.
- The arrows represent "IS_A" Relationships.
  - Bold arrows represent added Relationships resulting from generation of the short canonical form.

F.6.2 Simple computation based on the canonical form

The short canonical form reduces complexity and duplication in the defining characteristics without losing any of the information embedded in the definition. Thus it provides a minimal form of representation which can be used for comparing two *Concepts.*

Table F.1 compares the definitions of "Aluminum pedal bike" in the full and short canonical forms and illustrates that they are derivable from one another without any loss of information. It is a useful example to explain the canonical form.

Table F.1: Mapping between full and short canonical forms

Full form	*Derivation from Short Canonical form*
IS_A=Pedal bike	Derivable from "IS_A=Bike" + "Powered-by=Pedals"
IS_A=Aluminum machine	Derivable from "IS_A= Metal machine" + "Substance=Aluminum"
Origin=Man-made	Derivable from "IS_A= Metal machine" + definition of Metal machine
Made-of=Aluminum	Stated "Substance=Aluminum"
Moves-on=Two wheels	Derivable from "IS_A=Bike" + definition of "Bike"
Powered-by=Pedals	Stated "Powered-by=Pedals"
Short Canonical form	Mapping to Full form
IS_A=Bike	Derivable from "IS_A=Pedal bike" + definition of "Pedal bike"
IS_A= Metal machine	Derivable from "IS_A=Aluminum machine" + definition of "Aluminum machine"
Made-of=Aluminum	Stated "Substance=Aluminum"
Powered-by=Pedals	Stated "Powered-by=Pedals"

The significance of this is that the short canonical form is the simplest form that is sufficient to answer various questions about an "Aluminum pedal bike".

Example:

- Is an "aluminum pedal bike" a "machine"? – YES
- Is an "aluminum pedal bike" made of "metal" – YES
- Is an "aluminum pedal bike" a "motor bike"? – NO

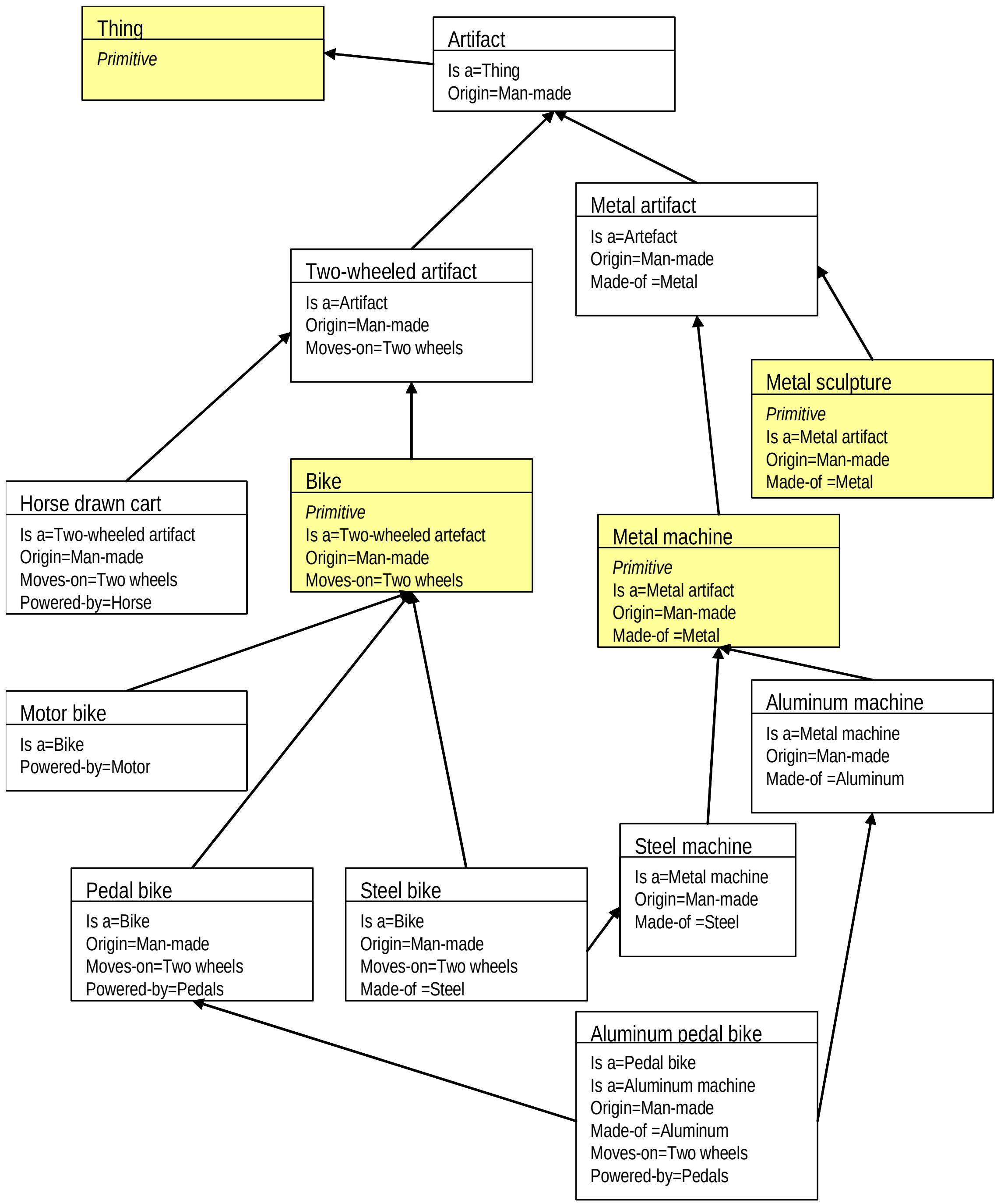


Figure F.1: Example of a set of defining relationships

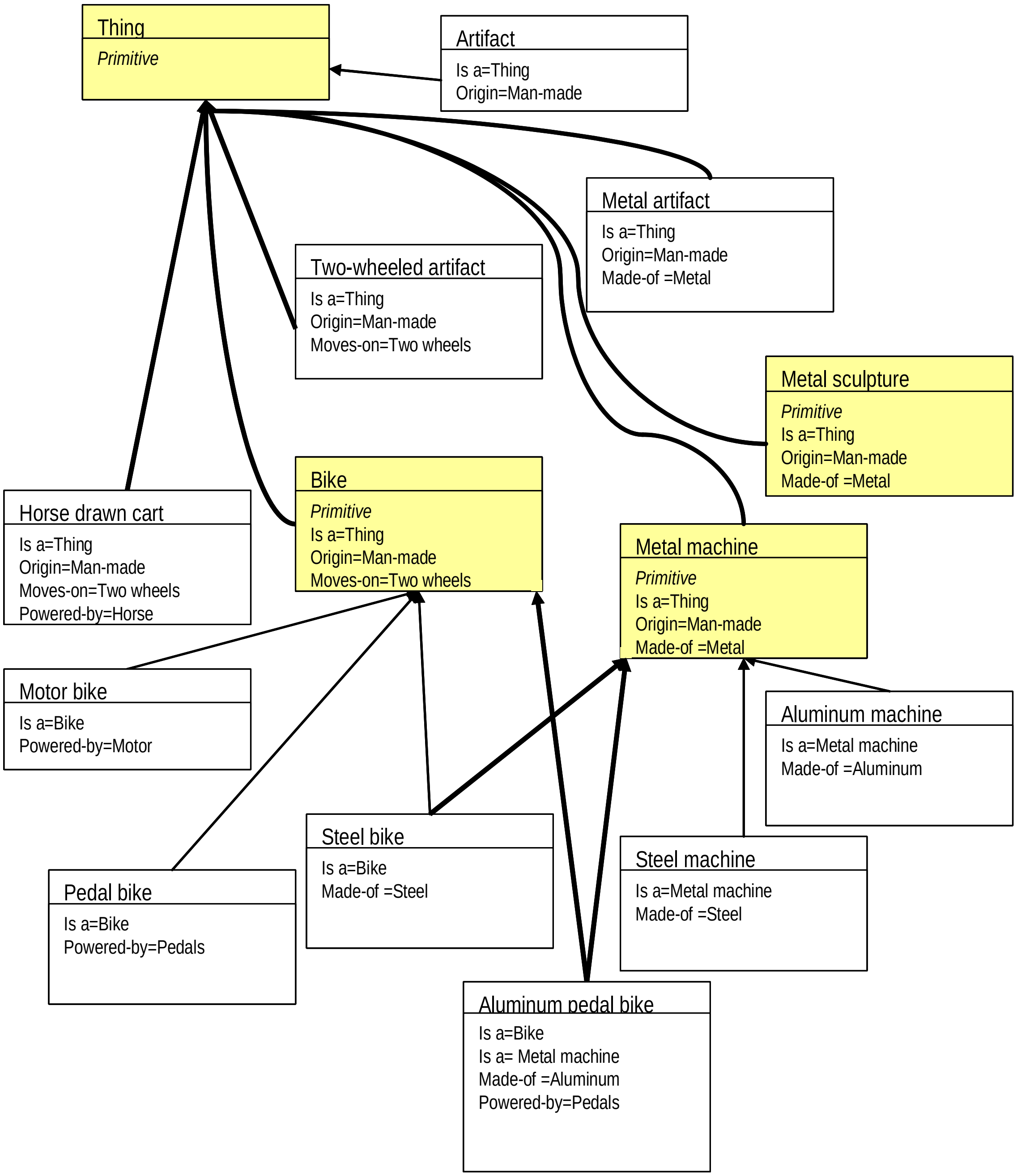


Figure F.2: Example of a short canonical form derived from previous figure

F.7 Comparing post-coordinated representation using the short canonical form

Post-coordinated representations of Concepts can also be compared with the short canonical form. However, primitive Concepts may impose limits on this. For example, the Concept "Steel pedal bike" does not exist in these illustrations. However, it could be represented by three different post-coordinated forms:

- "Pedal bike" + "Made-of=Steel"
- "Steel bike" + "Powered-by=Pedals"
- "Steel machine" + "Moves on=Two wheels" + "Powered-by=Pedals"

If the rules for generating the short canonical form are applied these representations the results are similar but incomplete representations of the expected short canonical form.

Table F.2: Equivalence of pre- and post-coordinated forms

Expected	Mapping from post-coordinated representation
Short Canonical form	"Pedal bike" + "Made-of=Steel"
IS_A=Bike	Derivable from "IS_A=Pedal bike" + definition of "Pedal bike"
IS_A=Metal machine	*Not derivable from this representation (see following notes)*
Made-of=Steel	Stated "Substance=Steel"
Powered-by=Pedals	Stated "Powered-by=Pedals"
Short Canonical form	"Steel bike" + "Powered-by=Pedals"
IS_A=Bike	Derivable from "IS_A=Steel bike" + definition of "Steel bike"
IS_A=Metal machine	*Not derivable from this representation (see following notes)*
Made-of=Steel	Derivable from "IS_A=Steel bike" + definition of "Steel bike"
Powered-by=Pedals	Stated "Powered-by=Pedals"
Short Canonical form	"Steel machine" + "Moves on=Two wheels" + "Powered-by=Pedals"
IS_A=Bike	*Not derivable from this representation (see following notes)*
IS_A=Metal machine	Derivable from "IS_A=Steel machine" + definition of "IS_A=Steel machine"
Made-of=Steel	Derivable from "IS_A=Steel machine" + definition of "IS_A=Steel machine"
Powered-by=Pedals	Stated "Powered-by=Pedals"
Moves-on=Two wheels	Stated "Moves-on=Two wheels"

Note:

This lack of complete equivalence illustrates the incompleteness of the definitions rather than any inherent limitation in the method. The enhanced definitions in the Figure F.3 and the resulting revised canonical form shown in Figure F.4 illustrate this.

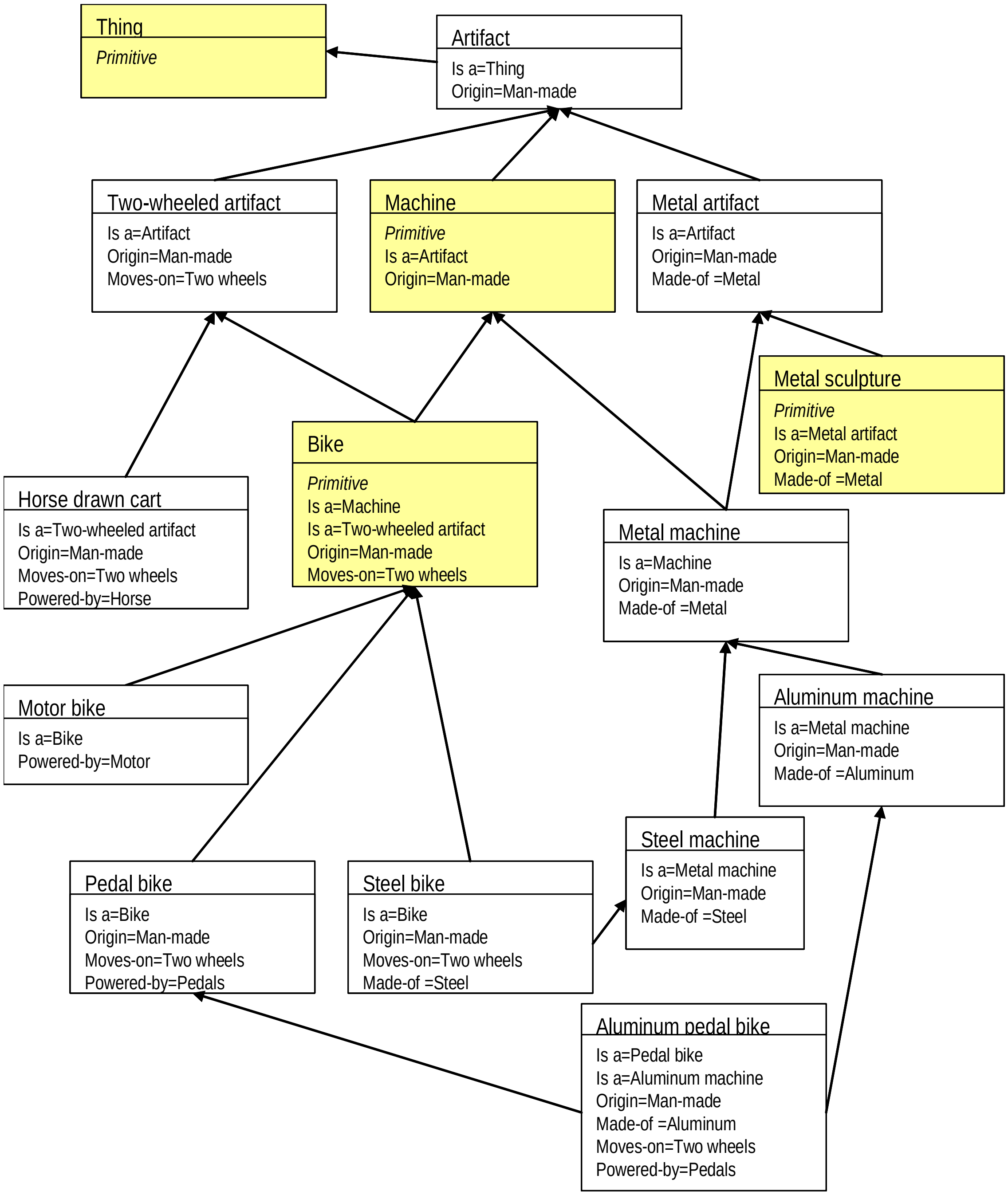


Figure F.3: Enhanced example of a set of relationships after adding a new concept (Machine)

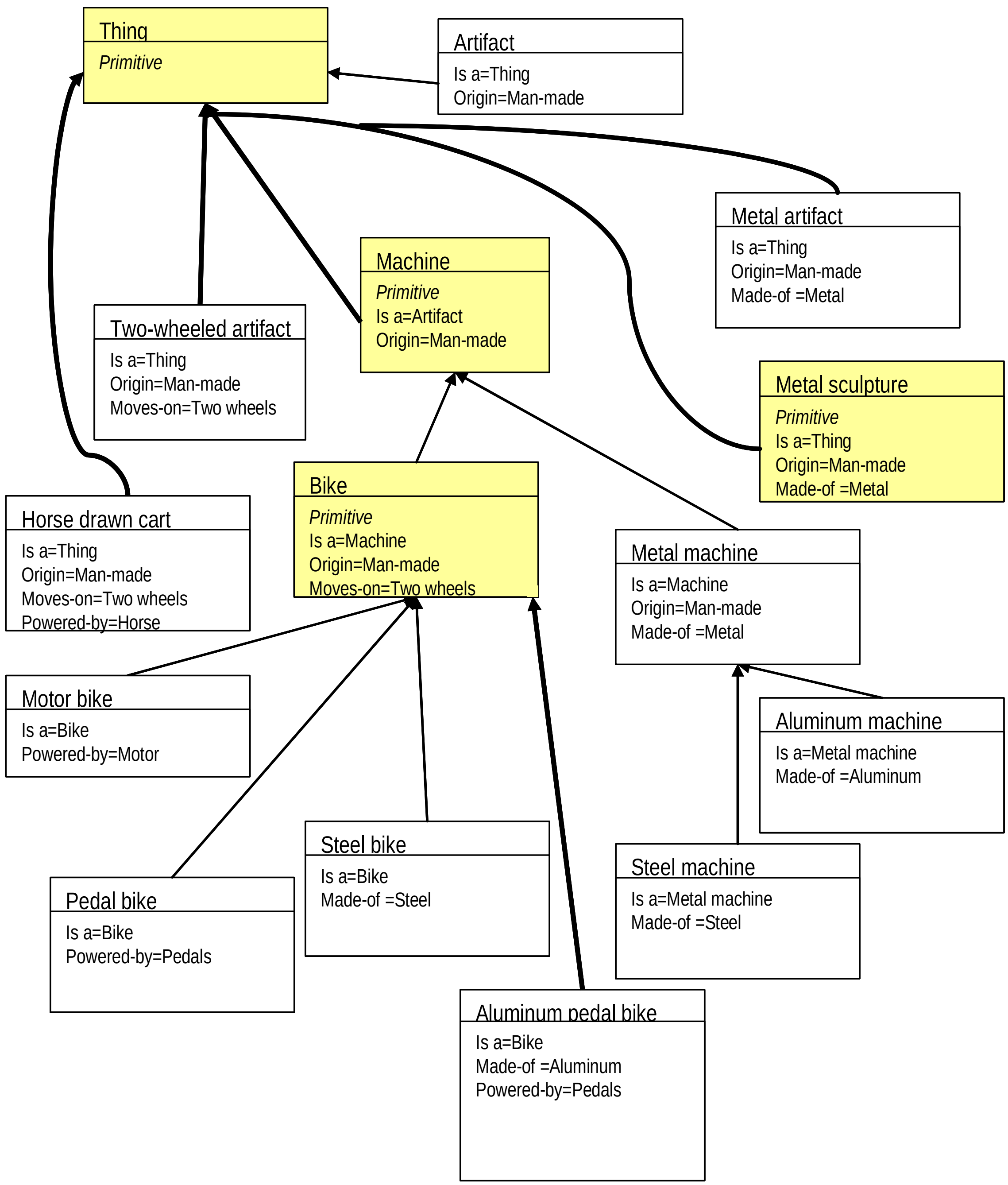


Figure F.4: Enhanced example of short canonical form after adding a new concept (Machine)

F.7.1 Improved comparisons with enhanced definitions

The changes between Figure F.3 and Figure F.4 are as follows

- Addition of a new primitive Concept "Machine"
- "Metal machine" is now fully defined as "IS_A=Machine"+ "Made-of=Metal"
- "Bike" remains primitive but has a new Relationship "IS_A=Machine".

As a result the short canonical form of "Aluminum pedal bike" as seen in Table F.3 is simplified "IS_A=Bike" + "Made-of=Aluminum" + "Powered-by=Pedals". The reference to "Metal machine" in the previous version is superfluous for the following reasons:

- "Metal machine" now has a primitive supertype "Machine"
- This is also PrimitiveSupertype of "Bike"
- "Bike" itself is a PrimitiveSupertype of the "Aluminum pedal bike"
  - Thus "Machine" is not a proximal PrimitiveSupertype.

Table F.3: Revised mapping between the full and short canonical forms

Full form	*Derivation from Short Canonical form*
IS_A=Pedal bike	Derivable from "IS_A=Bike" + "Powered-by=Pedals"
IS_A=Aluminum machine	Derivable from "IS_A= Bike" + "Bike – IS_A=Machine"+ "Substance=Aluminum" + Definition of "Aluminum machine"
Origin=Man-made	Derivable from "IS_A= Bike" + "Bike –IS_A=Machine" + Definition of "Machine"
Made-of=Aluminum	Stated "Substance=Aluminum"
Moves-on=Two wheels	Derivable from "IS_A =Bike" + definition of "Bike"
Powered-by=Pedals	Stated "Powered-by=Pedals"
Short Canonical form	Mapping to Full form
IS_A =Bike	Derivable from "IS_A=Pedal bike" + definition of "Pedal bike"
Made-of=Aluminum	Stated "Substance=Aluminum"
Powered-by=Pedals	Stated "Powered-by=Pedals"

Two of the post-coordinated representations discussed earlier now create canonical forms that are identical with the pre-coordinated representation.

- "Pedal bike" + "Made-of=Steel"
  - "IS_A=Bike" + "Made-of=Steel" + "Powered-by=Pedals"
- "Steel bike" + "Powered-by=Pedals"
  - "IS_A=Bike" + "Made-of=Steel" + "Powered-by=Pedals"

F.7.2 Residual limitations relating to primitive concepts

The third post-coordinated representation still produces a different result:

- "Steel machine" + "Moves on=Two wheels" + "Powered-by=Pedals"
  - "IS_A=Machine"+"Made-of=Steel"+"Powered-by=Pedals"+"Moves on=Two wheels"
- The difficulty is that "Bike" is stated to be primitive. This correctly recognizes that a machine, which moves on two wheels, may be something other than a "Bike". Thus this information provided is insufficient to specify that this is a "Bike". An examination of the possibilities might identify "Pedal bike" as the only Concept in the thesaurus

that “IS_A=Machine”+“Powered-by=Pedals”+“Moves on=Two wheels”. However, even this would not prove that this post-coordinated representation did not refer to some other type of device.

- In cases such as this, generating the canonical form correctly identifies the inherent ambiguity in some superficially similar post-coordinated representations.

Appendix G: Glossary

Active Concept	A Concept that is intended for active use. This is determined by the ConceptStatus. Concepts with status value "Current" (0) or "Limited" (6) are always regarded as active. Concepts with status "Pending Move" (11) are regarded as active if the Concept is not yet accessible in the new target Namespace. (see also Inactive Concept)
Active Description	A Description that is intended for active use. This is determined by a combination of the DescriptionStatus, the DescriptionType in the Language or Dialect in use and the ConceptStatus of the associated Concept. Descriptions are active when the following conditions apply: • Associated with an Active Concept, AND • DescriptionStatus "Current" (0), "Limited" (6) or "Pending Move" (8) AND • DescriptionType **not** unspecified (0) in a chosen Language/Dialect. (see also Inactive Description)
Attribute	Express characteristics of concepts. SNOMED CT concepts form relationships to other SNOMED CT concepts through attributes. ***Example:*** FINDING SITE All of the attributes used in modeling SNOMED CT concepts are themselves SNOMED CT concepts and can be found in the Linkage concept hierarchy.
Attribute-value pair	The combination of an attribute with a value that is appropriate for that attribute. ***Example:*** FINDING SITE = *Lung structure*
Axis Modification	An attribute-value pair (relationship) that does not result in a subtype of the Parent concept. ***Example:*** "*Family History of diabetes*" (with the added value "Family history") is not a subtype of the Clinical finding "*Diabetes*." These types of attributes shift the meaning of the concept, and the modified concept moves to a different SNOMED CT hierarchy. In the example, "*Family History of diabetes*" moves to the Situation with explicit context hierarchy.
Base Subset	A Subset used as a starting point the Subset Definition of another Subset.

Browser	A tool for exploring and searching the terminology content. A browser can display hierarchy sections and concept details (relationships between concepts, descriptions and Ids, etc.).
Canonical form	A table that contains the Canonical form expressions for all SNOMED CT Concepts. This table contains some (but not all) of the attributes present in the Relationships Table. It only contains the defining characteristics required to distinguish a Concept from its most proximate Primitive supertype Concepts. The set of "IS_A" subtype Relationships excludes Relationships to Fully defined (non-primitive) Concepts and includes instead appropriate additional Relationships to the most proximate supertypes.
Canonical Table	A table that contains the Canonical form expressions for all SNOMED CT Concepts. This table contains some (but not all) of the attributes present in the Relationships Table. It only contains the defining characteristics required to distinguish a Concept from its most proximate Primitive supertype Concepts. The set of "IS_A" subtype Relationships excludes Relationships to Fully defined (non-primitive) Concepts and includes instead appropriate additional Relationships to the most proximate supertypes.
Change Type	A field in the Component History Table that indicates the nature of a change to a Component made in a specified release of SNOMED CT.
Characteristic Type	A field in the Relationships Table that indicates whether a Relationship specifies a defining characteristic, a qualifying characteristic or a context-specific characteristic.
Check-digit	A digit used to check the validity of an SCTID. The check-digit is the final (right most) digit of the SCTID and it is calculated using the algorithm described in Appendix F of the Technical Reference Guide.
Child/Children	See Subtype.
Clinical Terms Version 3	See CTV3.

Component	An identifiable item in the main body of SNOMED CT, or in an authorized Extension. Each Component is a uniquely identifiable instance of one of the following: • Concept • Description • Relationship • Subset • Subset Member • Cross Map Set • Cross Map Target • History Component
Component History	A record of an addition or change in the status of a SNOMED CT Component in a particular Release Version. Each item of Component History is represented by a row in the Component History Table.
Component History Table	A data table consisting of rows each of which represents an item of Component History. The Component History Table is part of the History Mechanism.
ComponentID	A general term used to refer to the primary identifier of any SNOMED CT Component. ComponentIDs include ConceptIDs, DescriptionIDs, RelationshipIDs, SubsetIDs, CrossMapSetIDs, and TargetIDs. All ComponentIDs follow the form of the SCTID specification.
Concept	A clinical idea to which a unique ConceptID has been assigned. Each Concept is represented by a row in the Concepts Table.
Concept equivalence	Concept equivalence occurs when a post-coordinated expression has the same meaning as a pre-coordinated Concept or another post-coordinated expression.
ConceptID	A SNOMED Clinical Terms Identifier that uniquely identifies a Concept.
ConceptID1	A field in the Relationships Table and Canonical Table that refers to the first of two related Concepts. The first Concept is defined or qualified by a Relationship to the second Concept.
ConceptID2	A field in the Relationships Table and Canonical Table that refers to the second of two related Concepts. The second Concept defines or qualifies the first Concept.

Concepts Table	A data table consisting of rows, each of which represents a Concept. See also the specification of the Concepts Table.
ConceptStatus	A field in the Concepts Table that specifies whether a Concept is in current use. Values include “current,” “duplicate,” “erroneous,” “ambiguous,” and “limited.”
Context Attributes	These attributes define the context of a concept, such as the applicable time period, subject, and status (present/absent). These attributes all cause axis modification and, when used with Clinical findings or Procedures, shift the new concept to the Situation with explicit context hierarchy.
Context Concept Subset	A Subset used to specify the Concepts that form part of a context-domain.
Context Description Subset	A Subset used to specify the Descriptions that form part of a context-domain.
Context Domain	A context domain is a set of values that are, or may be, used in an identifiable logical setting in an application, protocol, query or communication specification. A context domain may be very broad (e.g. procedures or diagnoses) or very narrow (e.g. procedures performed by a specialty or possible values for a field in specific message).
ContextID	A field in the Subsets Table that specifies the context within which a Context Concept Subset or Context Description Subset is valid.
Context-specific characteristic	A Relationship to a target Concept that provides information about the source Concept that is true at a particular time or within a particular country or organization. Contrast with defining characteristic and qualifying characteristic. Referred to in CTV3 as a ‘Fact.’
Core	A SNOMED CT Component released by the IHTSDO.
Core Tables	See SNOMED CT Core Tables.
Cross Map	A reference from a Concept to a Cross Map Target. A Cross Map is part of a Cross Map Set. A Concept may have a single Cross Map or a set of alternative Cross Maps. Each Cross Map is represented as a row in the Cross Maps Table.

Cross Map Set	A set of Cross Maps that together provide a valid way of mapping some or all SNOMED CT Concepts to a specified Target Scheme. Alternative Cross Map Sets may exist for the same Target Scheme, if business rules or guidelines alter the appropriateness of particular mappings to that scheme. Each Cross Map Set is represented as a row in the Cross Map Sets Table.
Cross Map Sets Table	A data table consisting of rows each of which represents a Cross Map Set.
Cross Maps Table	A data table consisting of rows each of which represents a Cross Map.
Cross Map Target	A code or set of codes in a Target Scheme that together represent an appropriate mapping from a clinical statement expressed using SNOMED CT. Some Cross Map Targets may be derived from two or more associated statements and in these cases the combination can be expressed as a set of associated rules. Each Cross Map Target is represented as a row in the Cross Map Targets Table.
Cross Map Targets Table	A data table consisting of rows each of which represents a Cross Map Set.
CTV3 **Clinical Terms Version 3**	One of the source terminologies, along with SNOMED RT, used to develop SNOMED CT. CTV3 is UK Crown Copyright, distributed by the United Kingdom National Health Service Information Authority, and integrated into SNOMED CT. Also known as “Version 3 of the Read Codes.” See also Read Code.
CTV3ID	A field in the Concepts Table that contains the Clinical Terms Version 3 identifier (Read Code) for that Concept.
Data migration	Steps taken to enable legacy data to be accessible as part of a system that uses SNOMED CT. Options for data migration include actual conversion of the data or provision of methods for accessing the data in its original form.
Defining characteristic	A Relationship to a target Concept that is always true from any case of the source Concept. ***Example:*** “’*Topography (site)*’= ‘*Liver*’” is a defining characteristic of the Concept ‘*Liver biopsy*.’ Contrast with qualifying characteristic and context-specific characteristic. Referred to in CTV3 as an ‘Atom.’

Descendants	All subtypes of a concept, including subtypes of subtypes. ***Example:*** If a concept has four children, then descendants = those children plus all the concepts that are descended from those four children. See Subtype.
Description	A row in the Descriptions table. Each Description is assigned a unique DescriptionID and connects a Term and a Concept.
DescriptionID	A SNOMED CT Identifier that uniquely identifies a Description.
Descriptions Table	A data table consisting of rows, each of which represents a Description.
DescriptionStatus	A field in the Descriptions Table that specifies whether a Description is in current use. Values include “current,” “duplicate,” “erroneous,” “inappropriate,” and “limited.”
DescriptionType	A field in the Descriptions Table that specifies whether a Description is a Fully Specified Name, Preferred Term, or Synonym. The DescriptionType is language dependent and may be changed by applying a Language Subset. It may be “undefined” in the released Descriptions Table in which case the Description is not used unless an appropriate Language Subset is applied.
Dialect	A language modified by the vocabulary and grammatical conventions applied in a particular geographical or cultural environment.
Dualkey	A key used to facilitate textual searches of SNOMED CT that consists of the first three letters of a pair of words in a Description. All possible pairs of words in each Description may be paired irrespective of their relative position in the Description. Dualkeys are represented as a row in the Dualkeys Table.
Dualkey Table	A table in which each row represents a Dualkey.
Duplicate Term	A Term that occurs in several Active Descriptions. Duplicate terms are valid in SNOMED CT since the intention is to provide natural terms used by clinicians rather than to apply formalized phraseology. The formalized form is provided by the Fully Specified Name and these are not permitted to be duplicated. Although Duplicate Terms can be identified by string matching, a Duplicate Terms Subset is specified to indicate the presence and likely priority of duplicates when undertaking a search.
Duplicate Terms Subset	A Subset Type that identifies Duplicate Terms and allows a priority to be specified between these for use in searches.

Equivalence	See Word Equivalents, Phrase equivalence, and Concept equivalence.
Excluded Word	A word that in a given language is so frequently used, or has so poor a discriminating power, that it is suggested for exclusion from the indices used to support textual searches of SNOMED CT. Excluded Words are represented as a row in the Excluded Words Table.
Excluded Words Table	A data table in which each row represents an Excluded Word.
Expression	A collection of references to one or more concepts used to express an instance of a clinical idea. An expression containing a single concept identifier is referred to as a pre-coordinated expression. An expression that contains two or more concept identifiers is a post-coordinated expression. The concept identifiers within a post-coordinated expression are related to one another in accordance rules expressed in the SNOMED CT Concept Model. These rules allow concepts to be: • Combined to represent clinical ideas which are subtypes of all the referenced concepts - E.g. "tuberculosis" + "lung infection" • Applied as refinements to specified attributes of a more general concept. - E.g. "asthma" : "severity" = "severe" ***Notes:*** See "SNOMED CT Guide - Abstract Models and Representational Forms". This has been released as an External Draft in July 2005.
Extension	A data table or set of data tables that is created in accordance with the structures and authoring guidelines applicable to SNOMED CT but which may not be not edited, maintained and distributed by the IHTSDO. SNOMED CT Components are identified using SCTIDs, which are structured to ensure that Extensions are recognizable and can be traced to an authorized originator.
Extra-Relationship	A Relationship between two Concepts distributed as part of an Extension. An Extra-Relationship may relate SCT Concepts and/or Extra-Concepts.
Extra-Concept	A Concept distributed as part of an Extension.
Extra-Description	A Description which is distributed as part of an Extension. An Extra-Description may apply a Term to an SCT Concept or to an Extra-Concept.

Extra-Subset	A Subset distributed as part of an Extension. The members of an Extra-Subset may include SNOMED CT-Concepts, Extra-Concepts, SNOMED CT-Descriptions and Extra-Descriptions.
Fully defined	A Concept is fully defined if its defining characteristics are sufficient to differentiate a concept relative to its immediate supertype(s). A Concept which is not fully defined is Primitive. For example if the Concept "Red car" is defined as [is a=car] + [color=red] this is fully defined but the same definition applied to the Concept "Red sports car" is Primitive.
Fully defined concept	SNOMED CT concepts are either primitive or fully defined. A concept is primitive when its modeling (attributes and parents) does not fully express its meaning. Fully defined concepts can be differentiated from their parent and sibling concepts by virtue of their relationships. Primitive concepts do not have the unique relationships needed to distinguish them from their parent or sibling concepts.
Fully Specified Name	A phrase that describes a Concept uniquely and in a manner that is intended to be unambiguous. This phrase is in the FullySpecifiedName field of the Concepts Table and is also in the Descriptions Table.
Hierarchy	An ordered organization of concepts. General concepts are at the top of the hierarchy; at each level down the hierarchy, concepts become increasingly specialized. SNOMED CT concepts are arranged into Top-level hierarchies. Each of these hierarchies subdivides into smaller sub-hierarchies. Concepts are related by IS_A relationships to their more general parent concepts directly above them in a hierarchy. There is one concept from which the Top-level hierarchies descend called SNOMED CT concept or the "Root concept."
Historical Relationship	A Relationship that refers from an Inactive Concept to an Active Concept that duplicates, corrects, replaces or disambiguates it. Note that Historical Relationships are used in a way similar to References for Descriptions. However, as part of the Relationships Table they are more readily accessible for computation and retrieval of legacy data.
History Mechanism	A SNOMED CT distribution table that contains information about the history of changes to one of the core SNOMED CT tables. The History Mechanism is supported by two distribution tables: • Component History Table • References Table

Inactive Concept	A Concept that is not intended to be actively used. This is determined by the ConceptStatus. Concepts with status values other than "Current" (0), "Limited" (6) and "Pending Move" (11) are regarded as inactive. Concepts with status "Pending Move" (11) are regarded as inactive if the Concept is accessible in the new target Namespace as one of the release date. Inactive Concepts remain in SNOMED CT to support legacy data recorded when these Concepts were in active use. (see also Active Concept)
Inactive Description	A Description that is not intended to be actively used. This is determined by a combination of the DescriptionStatus, the DescriptionType in the Language or Dialect in use and the ConceptStatus of the associated Concept. Descriptions are inactive if one or more of the following apply: • Associated with an Inactive Concept, OR • DescriptionStatus not "Current" (0), "Limited" (6) or "Pending Move" (8) OR • DescriptionType unspecified (0) in a chosen Language/Dialect Inactive Descriptions remain in SNOMED CT to support legacy data recorded when these Descriptions were in active use. (see also Active Description)
InitialCapitalStatus	A field in the Descriptions Table that specifies whether the capitalization of the first character is significant. If the value of this field is "1" then the first character should remain either in upper or lower case as released. Otherwise the case of the first character may be changed to suit its full context in a sentence.
is a **IS_A** **"IS_A" relationship**	The Relationship Type that defines a supertype-subtype Relationship between two Concepts. Usually expressed as "subtype IS_A supertype." ***Example:*** Blister with infection IS_A infection of skin.
IsPrimitive	A field in the Concepts Table that indicates whether a Concept is Primitive or fully defined.
Keyword	A field containing potential search text in one of the WordKey Tables or a word excluded for key generation in the Excluded Words Table.

Kind-of-Value	The nature of a value that may be associated with a Concept. ***Example:*** The concept "*systolic blood pressure reading*" can label a numeric value. The Kind-of-Value that it labels is a pressure.
Language	A vocabulary and grammatical form that has been allocated an ISO639-1 language code. See also Dialect.
Language Subset	A Subset that specifies the status of various Terms according to a language or dialect and the DescriptionType for each Term.
LanguageCode	A field that indicates the Language and, optionally, Dialect applicable to a row in the Subsets Table, Descriptions Table or to an Excluded Words Table.
LinkedID	A field in the Subsets Table.
MapAdvice	A field in the Cross Maps Table that may contain human-readable advice on mapping.
MapConceptID	A field in the Cross Maps Table containing the identifier of the Concept that is the subject of the map.
MapOption	A field in the Cross Maps Table that specifies the order in which Cross Maps are tested for automated processing of cross mapping rules.
Mapping Mechanism	A mechanism for mapping to other terminologies and classifications. The Mapping Mechanism is supported by three distribution tables: • Cross Map Sets Table • Cross Maps Table • Cross Map Targets Table
MapPriority	A field in the Cross Maps Table that specifies which Cross Maps are most likely to apply to the associated Concept. The value "0" indicates a default map.
MapRule	A field in the Cross Maps Table that may contain a computer-processable representation of a rule that determines when this map should be used.
MapSetID	A SNOMED CT Identifier that uniquely identifies a Cross Map Set.
MapSetName	A field in the Cross Map Sets Table that names that Cross Map Set.
MapSetRealmID	A field in the Cross Map Sets Table that indicates the Realm in which a set of Cross Maps is applicable.

MapSetRuleType	A field in the Cross Map Sets Table that indicates whether any computer-processable rules are present in the associated Cross Maps or Cross Map Targets and, if so, what form of expression is used to represent these rules.
MapSetSchemeID	A field in the Cross Map Sets Table, which identifies the classification or coding scheme that is the target of a Cross Map Set.
MapSetSchemeName	A field in the Cross Map Sets Table that contains the plain text name of the classification or coding scheme that is the target of a Cross Map Set.
MapSetSchemeVersion	A field in the Cross Map Sets Table that identifies the version of the classification or coding scheme that is the target of a Cross Map Set.
MapSetSeparator	A field in the Cross Map Sets Table that contains a character that acts as a separator between target codes in the Cross Map Targets Table.
MapSetType	A field in the Cross Map Sets Table that indicates whether the Cross Maps associated with this Cross Map Set are all simple one to one maps or include, one to many and/or choices of alternative maps.
MapTargetID	A field in the Cross Maps Table, which refers to the TargetID of a row in the Cross Map Targets Table that contains a target scheme mapping for a specified Concept.
MemberID	A field in the Subset Members Table, which refers to the ComponentID of the Concept, Description or other Component that is a member of a specified Subset.
MemberStatus	A field in the Subset Members Table, which indicates the inclusion, priority or order of an identified Subset Member in a specified Subset.
Migration	See Operational migration, Data migration, and Predicate migration.
Modeler	The SNOMED CT content developers who model the terminology. Also referred to as “Clinical Editors” or “Authors.”
Modeling	The process of editing concepts to reflect their unique definition and meaning.
Moved Elsewhere	A Status value applicable to a Component that has been moved to another Namespace. Concepts or Descriptions may be moved from an Extension to the SNOMED CT core, from the core to an Extension or between one Extension and another. Moves occur if responsibility for supporting the Concepts changes to another organization.

Namespace **Namespace-identifier**	A block of identifiers allocated for use by an organization creating Extensions to SNOMED CT. The Namespace identifier is part of the SCTID and its structure is described in Appendix C of the Technical Reference Guide.
Namespace Concept	A Concept that exists to represent a SNOMED CT Namespace identifier. All Namespace Concepts are direct subtypes of the Concept "Namespace Concept" which is a subtype of the Top-level Concept "Special Concept." Namespace Concepts are used as the target of Historical Relationships and References when a Component is moved from one Namespace to another.
Navigation	The process of locating a Concept by traversing Relationships or Navigational links. For example, moving from a supertype Concept to more refined Concepts, from a specific Concept to a more general Concept or from a Concept to its Defining characteristics. Navigational links allow navigation to follow intuitive routes through SNOMED CT even where there are no direct supertype or subtype Relationships.
Navigation Concept	A Concept that exists only to support Navigation. A Navigation Concept is not suitable for recording or aggregating information. All Navigation Concepts: • Are direct subtypes of the concept "Navigational Concept" • Have no other supertype or subtype Relationships • Are linked to other Concepts only by Navigational links
Navigation Link	An association between two Concepts that supports Navigation between Concepts. Navigation Links generate a hierarchy which has three distinct differences from the subtype hierarchy defined by "IS_A" Relationships this hierarchy: • Does not effect the semantic definitions of Concepts; • Specifies a display order Concepts within a set of Concepts linked to a common parent. Navigation Links are distributed as a Navigation Subset. Alternative Navigation Subsets may be specified and applied to vary the navigational hierarchy to meet the needs of particular groups of users.
Navigation Subset	A Subset that specifies sets of Navigation Links between Concepts.
Operational migration	Steps taken to enable an organization that either used a previous coding scheme (or no clinical coding scheme) to make use of SNOMED CT.

Partition-identifier	A pair of digits that indicate whether an SCTID identifies a Concept, Description, Relationship, Subset, History or Extension component. The partition-identifier consists of the second and third digits from the right of the SCTID.
Pending Move	A Status value applicable to a Component that is thought to belong in a different Namespace but which is maintained with its current SCTID while awaiting addition to the new Namespace. A new Concept and associated Description may be added with this Status where a missing SNOMED CT Concept is urgently required to support the needs of a particular Extension. Existing Concepts can also use this status when it is recognized that they should be moved to another namespace (organization) or to the core namespace. See also Moved Elsewhere.
Phrase equivalence	Two words or phrases with a similar meaning. ***Example:*** "renal calculus" and "kidney stone." See Word Equivalents.
Post-coordination	Representation of a clinical idea using a combination of two or more concept identifiers. A combination of concept identifiers used to represent a single clinical idea is referred to as a post-coordinated expression (see expression). Many clinical ideas can also be represented using a single SNOMED CT concept identifier (see pre-coordination). Some clinical ideas may be represented in several different ways. SNOMED CT technical specifications include guidance of logical transformations that reduce equivalent expressions to a common canonical form. ***Example:*** SNOMED CT includes the following concepts: • Fracture of bone (ConceptID= 125605004) • Finding site (ConceptID= 363698007) • Bone structure of femur (ConceptID= 181255000) SNOMED CT also includes a pre-coordinated concept for this procedure. • Fracture of femur (ConceptID= 71620000) It is possible to represent "fracture of femur" in different ways: • 71620000 (pre-coordinated expression) • 125605004 : 363698007 = 181255000 (post-coordinated expression).

Pre-coordination	Representation of a clinical idea using a single concept identifier. A single concept identifier used to represent a specific meaning is referred to as a pre-coordinated expression (see expression). SNOMED CT also allows the use of post-coordinated expressions (see post-coordination) to represent a meaning using a combination of two or more concept identifiers. However, including commonly used concepts in a pre-coordinated form makes the terminology easier to use. For examples, see post-coordination.
Predicate migration	Steps taken to enable pre-existing data retrieval predicates (including queries, standard reports and decision support protocols) to be converted or utilized in a system using SNOMED CT.
Preferred Term	The Term that is deemed to be the most clinically appropriate way of expressing a Concept in a clinical record. Preferred Term is one of the three types of terms that can be indicated by the DescriptionType field.
Primitive Concept	A Concept is Primitive if its defining characteristics are insufficient to define it relative to its immediate supertype(s). For example, if the Concept “Red sports car” is defined as [is a=car] + [color=red] this is Primitive but the same definition applied to the Concept “Red car” is Fully defined.
Qualifying attribute	Some SNOMED CT concepts can have qualifying attributes, which are optional non-defining relationships that may be applied by a user or implementer in post-coordination. The qualifier value mechanism in SNOMED CT constrains the possible values an implementer can select in assigning a qualifying characteristic to a concept. Example: “*Acute bronchitis*” can be post-coordinated using the pre-coordinated SNOMED CT concept *Bronchitis (disorder)* and adding the qualifying attribute COURSE and the value *Acute (qualifier value)*.
Qualifying characteristic	A Relationship to a target Concept that specifies a possible qualification of the source Concept. ***Example:*** “’Revision status’ = ‘Conversion from other type of arthroplasty’” is a possible qualifying characteristic of ‘*Hip replacement*.’ Contrast with defining characteristic and context specific characteristic. Referred to in CTV3 as a ‘Qualifier.’

Read Code **CTV3ID**	A five-character code allocated to a concept of term in CTV3. Note that codes allocated in Read Codes Version 2 and the Read Codes 4-Byte Set are also included in CTV3. In the case of 4-byte codes the original code is prefixed by a full stop ('.').
Read Codes 4-Byte Set	The first version of the clinical coding scheme developed by Dr. James Read. The Read Codes 4-Byte Set is UK Crown Copyright distributed by NHS IA, and integrated into SNOMED CT.
Read Codes Version 2	The second version of the clinical coding scheme developed by Dr. James Read. Read Codes Version 2 is UK Crown Copyright distributed by NHS IA, and integrated into SNOMED CT.
Realm	A sphere of authority, expertise, or preference that influences the range of Concepts and Descriptions required, or the frequency with which they are used. A Realm may be a nation, an organization, a professional discipline, a specialty, or an individual user. See also Realm Concept Subset.
Realm Concept Subset	A Subset of Concepts applicable to a particular Realm.
Realm Description Subset	A Subset of Descriptions applicable to a particular Realm.
Realm Relationship Subset	A Subset of Relationships applicable to a particular Realm. (for future use)
RealmID	A field in the Subsets Table that identifies the Realm within which the specified Subset is applicable.
Reason	A field in the Component History Table which provides a text description of the reason for a change in the Status of a Component.
Reference	An association between a non-current SNOMED CT Component and a current Component that duplicates it, replaces it, or is related to it. Each Reference is represented by a row in the References Table.
ReferencedID	A field in the References Table that identifies the current Component which replaces it or is duplicated by a non-current Component.
References Table	A data table consisting of rows each of which represents a Reference. The References Table is part of the History Mechanism.
ReferenceType	A field in the References Table that indicates whether a specified non-current Component was replaced by, duplicated by, similar to or an alternative form of the referenced current Component.

Refinability	A field in the Relationships Table, which indicates whether it is permissible to refine the value of a defining characteristic or qualifying characteristic to represent a more refined Concept.
Relationship	An association between two Concepts (each identified by a ConceptID). The nature of the association is indicated by a RelationshipType. Each Relationship is represented by a row in the Relationships Table.
Relationship group	Relationships for concepts that are logically associated with each other. The Relationship group field in the Relationships Table is used to group these rows together for a concept. For example, where a particular type of prosthesis is inserted in a joint the defining characteristics describing the prosthesis type would be in one group whereas those describing the location or laterality of the joint would be in another group.
Relationship Type **Type of Relationship**	The nature of a Relationship between two Concepts. Relationship Types are represented in SNOMED CT by Concepts. The RelationshipType field indicates the nature of the Relationship by referring to the appropriate ConceptID.
RelationshipID	A SNOMED Clinical Terms Identifier that uniquely identifies a Relationship.
Relationships Table	A data table consisting of rows, each of which represents a Relationship.
Release Version	An identifiable set of SNOMED CT tables distributed on or after a particular date for use in SCT Enabled Applications. Each Release Version is referred to by the ISO format date of which this set of files was distributed (or was scheduled for distribution). ***Example:*** Release Version “20030131” refers to the version released on January 31, 2003.
ReleaseVersion	A field in the Component History Table which indicates the SNOMED CT release in which a Component was added or changed.
Retired Concept	Concepts that are no longer considered current are called “non current” or “inactive” rather than “retired.”
Role	Another name for Attribute. See Attribute.
Role Hierarchy	A role hierarchy is a hierarchical relationship between the Relationship Types themselves. Concepts use roles (relationships) at any level that is appropriate. Data queries can be developed that take advantage of the role hierarchies.

Root Concept	A single special Concept that represents the root of the entire content of SNOMED CT. All other Concepts are related to this Concept via at least one series of Relationships of the Relationship Type "IS_A" (i.e. all other Concepts are regarded as subtypes of the Root Concept).
SCT Enabled Application	A software application designed to support the use of SNOMED CT.
SCT-Concept **SCT-Description** **SCT-Relationship** **SCT-Subset**	A Concept, Description, Relationship, or Subset distributed by IHTSDO as part of SNOMED CT. The prefix "SCT-" is used only where it is necessary to distinguish clearly between elements distributed by IHTSDO and Extensions (denoted by the prefix "Extra-"). For further information refer to the glossary entries for the unprefixed names.
SNOMED®	The **S**ystematized **No**menclature of **Med**icine.
SNOMED Clinical Terms **SNOMED CT®**	The clinical terminology maintained and distributed by the IHTSDO. SNOMED Clinical Terms is the result of the merger of the CTV3 and SNOMED RT.
SNOMED Clinical Terms Identifier **SNOMED CT Identifier** **SCTID**	A unique identifier applied to each Concept, Description, or Subset. The SCTID includes a check-digit and a partition-digit.
SNOMED CT Derivative	Documentation, subsets, cross-mappings, extensions, and other files.
SNOMED CT Core Tables	Refers to the SNOMED CT Concept, Relationship, and Description Tables.
SNOMED Reference Terminology **SNOMED RT**	The latest version of SNOMED® prior to the collaborative development to develop SNOMED Clinical Terms. One of the source terminologies, along with CTV3, used to develop SNOMED CT. SNOMED RT was sponsored by the College of American Pathologists (CAP).
SNOMEDID	A field in the Concepts Table that contains the SNOMED RT identifier for the Concept.
Soft default	A "soft default" is the set of assumptions that can be made if a concept appears in a record with no explicit context in its definition. Procedures and clinical findings have these soft defaults: • Present • In the patient

Status	The Status of a Component indicates whether it is in current use, and, if not, provides a general indication of the reason that it is not recommended for current use. The Status of a Concept is referred to as ConceptStatus and the Status of a Description is referred to as DescriptionStatus.
Subset	Subsets represent groups of Components that share specified characteristics that affect the way they are displayed or otherwise accessible within a particular realm, specialty, application or context.
Subset Definition	A series of clauses that specifies the nature of a Subset and determines its membership. • A Subset Definition is an alternative form of representation applicable to many Subsets. See also Subset Definition File.
Subset Definition Clause	A statement that refers directly or indirectly to one or more Concepts, Descriptions or Relationships and indicates whether the reference item(s) should be removed from or added to a specified Subset. In the case of additions to a Subset the clause also specifies the Member Status to be assigned to those additions.
Subset Definition File	A file containing a series of clauses that define the nature and membership of a Subset. The Subset Definition File is an XML document that contains the Subset Definition for a single Subset. A Subset Definition File can be used to generate the appropriate Subsets Table row and Subset Members Table rows to represent a Subset. Therefore applications need not support this alternative form of Subset representation.
Subset Member	A Concept, Description, Relationship or another Subset that is part of a specified Subset.
Subset Members Table	A data table consisting of rows each of which refers to a single Subset Member.
Subset Type **Type of Subset**	An indication of the type of component that may be a member of a Subset.
Subset Version	A version number assigned to particular release of a Subset.
SubsetID	A SNOMED CT Identifier that uniquely identifies a Subset.
SubsetName	A field in the Subsets Table that contains a human readable name for the Subset.

SubsetOriginalID	A field in the Subsets Table that identifies the first version of the Subset on which this Subset is based. For the first version of a Subset the SubsetOriginalID and Subset fields contain the same value. For each subsequent version the Subset Version is incremented and a new SubsetID is allocated but the SubsetOriginalID field retains the same value in all versions.
Subsets Table	A data table consisting of rows each of which represents a Subset.
Subtype	A specialization of a concept, sharing all the definitional attributes of the parent concept, with additional granularity. ***Examples:*** *Bacterial infectious disease* is a subtype of *infectious disease*. *Bacterial septicemia*, *bacteremia*, *bacterial peritonitis*, etc., are subtypes of *bacterial infectious disease* (and *infectious disease* as well). In short, the concepts in a hierarchy that are directly related to a parent concept via the "IS_A" relationship. Distinguished from "Descendants" which explicitly includes subtypes of subtypes. Subtype is usually intended to refer to only the concepts that are immediately under the subject concept – that is, one level down in the hierarchy.
Synonym	A Term which is an acceptable alternative to the Preferred Term as a way of expressing a Concept.
Target Code	A code or other identifier within a Target Scheme.
Target Scheme	A terminology, coding scheme or classification to which some or all SNOMED CT Concepts are cross-mapped. Mappings from SNOMED CT to a Target Scheme are represented by one or more Cross Map Sets.
TargetAdvice	A field in the Cross Map Targets Table that may contain human readable advice on the circumstances in which this Cross Map Target is applicable.
TargetCodes	A field in the Cross Map Targets Table that contains one or more codes or identifiers in the target scheme or classification. If there is more than one Target Code they are separated by a separator specified in the associated row of the Cross Map Sets Table.
TargetID	A SNOMED Clinical Terms Identifier that uniquely identifies a Cross Map Target.
TargetRule	A field in the Cross Map Targets Table that may contain computer-processable rules specifying circumstances in which this Cross Map Target is applicable.
TargetSchemeID	A field in the Cross Map Targets Table that identifies the target coding scheme or classification to which this Cross Map Target applies.

Term	A text string represents the Concept. The Term is part of the Description. There are multiple descriptions per Concept.
Terminology server	Software that provides access to SNOMED CT (and/or to other terminologies). A Terminology server typically supports searches and Navigation through Concepts. A server may provide a user interface (e.g. a browser or set of screen controls) or may provide low-level software services to support access to the terminology by other applications. See Section 6.
Top-Level Concept	A Concept that is directly related to the Root Concept by a single Relationship of the Relationship Type “IS_A.” All other Concepts are descended from one Top-Level Concept via at least one series of Relationships of the Relationship Type “IS_A” (i.e. all other Concepts are subtypes of one Top-level Concept).
Unicode	A standard character set, which represents most of the characters used in the world using a 16-bit encoding. Unicode can be encoded in using UTF-8 to more efficiently store the most common ASCII characters.
UTF-16	A standard method of directly encoding Unicode using two bytes for every character. See also UTF-8.i.e.
UTF-8	A standard method of encoding Unicode characters in a way optimized for the ASCII character set. UTF-8 is described Appendix E of the TRG.
Word Equivalent	A word or abbreviation that is stated to be equivalent to one or more other words, phrases or abbreviations for the purposes of textual searches of SNOMED CT. Word Equivalents and Phrase Equivalents are represented as rows in the Word Equivalents Table.
Word Equivalents Table	A data table in which each row represents a Word Equivalent.
WordBlockNumber	A field in the Word Equivalents Table, which links together several rows which have an identical or similar meaning.
WordKey Table	A data table relating each word used in SNOMED CT (other than Excluded Words) to the Descriptions.
WordRole	A field in the Word Equivalents Table, which specifies the usual usage of this word, abbreviation or phrase, or the usage in which it has a similar meaning to the text in one or more rows of the table that share a common WordBlockNumber.

WordText	A field in the Word Equivalents Table, which contains a word, phrase, acronym or abbreviation that is considered to be similar in meaning to the text on one or more other rows of the table that share a common WordBlockNumber.
WordType	A field in the Word Equivalents Table, which specifies whether this row contains a word, phrase, acronym or abbreviation.